TAMPERE UNIVERSITY OF TECHNOLOGY

SUTANSHU SAKSENA RAJ

FOVEATED NON-LOCAL MEANS DENOISING FOR COLOR IMAGES, WITH CROSS-CHANNEL PARADIGM

Master of Science Thesis

Examiner: Dr. Alessandro Foi

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Master's Degree Programme in Information Technology
SUTANSHU SAKSENA RAJ : Foveated Non-Local Means Denoising of Color Images, with Cross-Channel Paradigm
Master of Science Thesis, 69 pages
August 2016
Major: Signal Processing
Examiner: Dr. Alessandro Foi
Keywords: Color Image Modeling, Foveation, Non-local Similarity, HVS, Denoising.

Foveation, a peculiarity of the HVS, is characterized by a sharp image having maximal acuity at the central part of the retina, the fovea. The acuity rapidly decreases towards the periphery of the visual field. Foveated imaging was recently investigated for the purpose of image denoising in the Foveated Non-local Means (FNLM) algorithm, and it was shown that for natural images the foveated self-similarity is a far more effective regularization prior than the conventional windowed self-similarity. Color images exhibit spectral redundancy across the R, G and B channels which can be exploited to reduce the effects of noise.

We extend the FNLM algorithm to the removal of additive white Gaussian noise from color images. The proposed Color-mixed Foveated NL-means algorithm, denominated as C-FNLM, implements the concept of foveated self-similarity, along with a cross-channel paradigm to exploit the correlation between color channels. The patch similarity is measured through an updated foveated distance for color images. In C-FNLM, we derive the explicit construction of an unified operator which explores the spatially variant nature of color perception in the HVS.

We develop a framework for designing the linear operator that simultaneously performs foveation and color mixing. Within this framework, we construct several parametrized families of the color-mixing operation. Our analysis shows that the color-mixed foveation is a far more effective regularity assumption than the windowing conventionally used in NL-means, especially for color image denoising where substantial improvement was observed in terms of contrast and sharpness. Moreover, the unified operator is introduced at a negligible cost in terms of the computational complexity.

ACKNOWLEDGEMENTS

As my time being a Masters graduate draws near, I'd like to take this opportunity to thank my advisor, Dr. Alessandro Foi, whose guidance, friendly-nature and creativity have been a source of inspiration for me. I believe that my enlightening discussions with Alessandro have played a crucial role in defining my fondness for images, and I feel grateful to him for sharing his thoughts with me. This thesis could not have taken a coherent shape without his patience and support.

I feel privileged to have had so many competent professors who nurtured my love and interest for signals. Additionally, I owe special thanks to Dr. Giacomo Boracchi, whose insight made possible the construction of certain ideas in this thesis. Also, I would be remiss without mentioning Florin Ghido, whose vision and enthusiasm have been a guiding example for me.

A special shout-out to all those friends who, due to circumstances, are no longer a part of my life. I thank you all for sharing your knowledge with me, for the impassioned discussions we engaged in, and for your patience with all my flaws. I equally appreciate the endurance of all those friends who are still an integral part of my life and with whom I have shared memorable experiences. Your outlook towards life has influenced my perspective and your company is always a pleasure.

Finally, I can never forget the support and sacrifice of my parents and my sister. You have provided me with immense love, humour and an endless desire to indulge in *bakchodi*. Also, I have a great set of cousins who have filled with joy all our interactions; not to forget, their "joint effort" in all my shenanigans. One sentence cannot justify the way you all have moulded my life.

I am an amalgamation of the best and worst in all of you.

Sutanshu Saksena Raj

CONTENTS

NOTATIONS

CCD Charge-coupled Device

CMOS Complementary Metal-oxide Semiconductor

AWGN Additive white Gaussian Noise

CLT Central Limit Theorem

i.i.d Independent and Identically Distributed

PDF Probability Density Function

MSE Mean Square Error

PSNR Peak Signal-to-noise Ratio

SSIM Structural Similarity Index

BM3D Block-matching and 3-D Filtering

DCT Discrete Cosine Transform

DWT Discrete Wavelet Transform

STFT Short-time Fourier Transform

DFT Discrete Fourier Transform

DDID Dual-domain Image Denoising

HVS Human Visual System

CFF Critical Fusion Frequency

DoG Difference of Gaussians

FREAK Fast Retina Keypoint

SIFT Scale-invariant Feature Transform

PSF Point-spread Function

E Mathematical Expectation

FNLM Foveated Non-local Means

CM Color-mixing

C-FNLM Color-Foveated Non-local Means

LIST OF FIGURES

1. INTRODUCTION: IMAGE DENOISING

In this thesis, a digital image is considered to be a two-dimensional function y, which is defined as:

$$y : X \to \mathbb{R}, \text{ where } X \subset \mathbb{Z}^2 \tag{1.1}$$

where $x \in X \subset \mathbb{Z}^2$ is a two-dimensional spatial coordinate called *pixels* in the domain X, and $y(x)$ represents the intensity value [1] of the grayscale image y at the position indexed by variable x. Thus, an image is a 2-D array whose elements are pixel values.

In certain applications, the image capturing process is the result of light intensity measurements made by CCD or CMOS sensors. The incident light (photons) impinging upon each sensor element is transformed into electrons, then converted to electrical voltage, before undergoing pre-processing in the digital form. This form undergoes specific adjustments, such as interpolation, gamma correction and color tone-mapping, before the final image is constructed.

1.1 Noise Models

Noise is an undesirable random component in an observed signal which corrupts the signal acquisition process. Images captured by digital imaging sensors typically contain noise. These occur due to the uncertain nature of photon emission and sensing, i.e. even for a light source with constant intensity, the number of photons striking the sensors is not fixed during a constant time interval, resulting in photon-counting noise (or, shot noise). Noise corrupting the image is introduced in different forms at various stages of the image formation, and can be divided into two main categories: signal-dependent noise, and signal-independent noise. The signal-dependent noise is essentially due to the photon-counting process, whereas the signal-independent noise is due to electric and thermal noise.

Given the unknown original image $y : X \to \mathbb{R}$, its noisy observations $z(x)$ are:

$$z(x) = y(x) + \eta(x) \qquad x \in X \subset \mathbb{Z}^2 \tag{1.2}$$

[1] For color images, $y(x)$ represents a triplet of values corresponding to the (R,G,B) color components; whereas for grayscale images, $y(x)$ is the observed intensity value.

where $z : X \to \mathbb{R}$ is the observed noisy image, and $\eta : X \to \mathbb{R}$ is assumed to be the noise corrupting the signal at every pixel. The specific measurement method determines the type of noise model. Generally, its behaviour is most often described by using random variables following either a Gaussian or a Poisson distribution, or a combination of both.

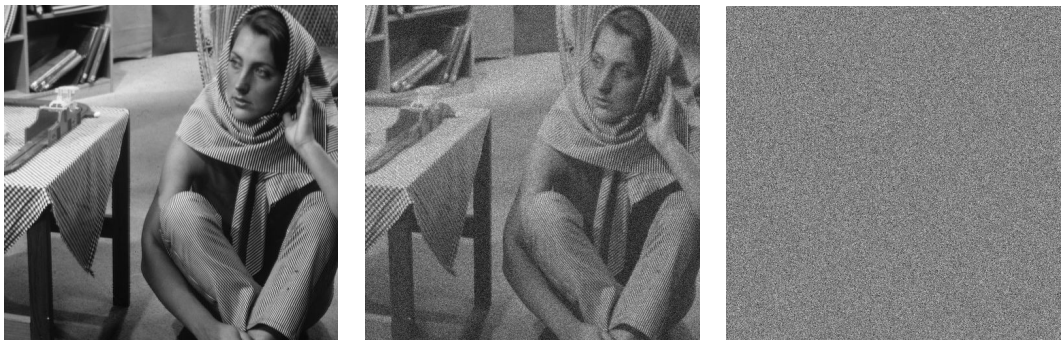
1.1.1 Gaussian Model

The most common probabilistic model used to approximate the effect of noise in corrupted images is the additive white Gaussian noise (AWGN). It is independent and identically distributed (i.i.d.), signal-independent and normally distributed.

The i.i.d. condition dictates that the variance σ^2 of the noise component η is constant over the image, and that the noise samples are drawn independently of each other. With reference to Eq. (1.2), the white Gaussian noise is

$$\eta(\,\cdot\,) \sim \mathcal{N}(0\,,\,\sigma^2)\ , \tag{1.3}$$

which is not an accurate representation for digital imaging devices as it ignores, among other factors, signal-dependent shot noise. Regarding the probability distribution of the noise, the assumption of Gaussianity is a direct consequence of the Central Limit Theorem (CLT): when an image is well-exposed, i.e. a large number of photons impinge upon the imaging sensors, the probability density function (PDF) of the noise closely resembles a Gaussian PDF.



Original Image Noisy Image Gaussian Noise

Figure 1.1: Gaussian noise realization of a two-dimensional image.

1.1.2 Poisson Model

Given the discrete nature of light, natural images are necessarily influenced by shot noise. The number of photons impinging upon a photon-counting device, such as those used for medical and astronomical imaging, are modeled as a Poisson

distribution. A simplifying assumption is that the degrading effects of all signal-independent sources are insignificant compared to signal-dependent noise.

Formally, each observation $z(x)$ of a noisy image z is defined as an independent random variable taken from a Poisson distribution with parameter proportional to the original image $y(x)$:

$$z(x) \sim \mathcal{P}(\chi \cdot y(x)) \tag{1.4}$$

where $x \in X \subset \mathbb{Z}^2$ and χ is a positive real number. It should be noted that both the expected value $\mathbb{E}$ and the variance $\mathbb{V}$ of the image z is the underlying intensity value to be estimated [56]:

$$\mathbb{E}[z(x)] = \mathbb{V}\text{ar}[z(x)] = \chi \cdot y(x) \qquad x \in X$$

Therefore, the noise η can be formally defined as:

$$\eta(x) = z(x) - \mathbb{E}[z(x)] \tag{1.5}$$

The standard deviation of a Poisson distribution is equal to the square root of its mean. Due to this, the relative effect of Poisson noise increases (i.e. the signal-to-noise ratio decreases) as the intensity value decreases.

Original Image Noisy Image Poisson Noise

Figure 1.2: Poisson noise realization of a two-dimensional image.

The expected number of photons impinging upon a sensor per unit time interval is proportional to the incident scene irradiance. An interesting property of the Poisson distribution is that for sufficiently high [2] values of the parameter λ, the Poisson distribution can be approximated as a Gaussian distribution with both mean and variance equal to λ :

$$\mathcal{P}(\lambda) \approx \mathcal{N}(\lambda, \lambda) \tag{1.6}$$

[2] E.g. $\lambda > 20$.

1.1.3 Poisson-Gaussian Model

The Poisson-Gaussian model describes the statistical behaviour of noise corrupting the unprocessed (or, raw) data generated by cameras. As the name suggests, it is composed of both the signal-dependent Poissonian and signal-independent Gaussian noise sources. The model comprises of both a multiplicative scaled Poisson term, and an additive Gaussian term.

We express the noisy image z to include a generic signal-dependent noise model, at pixel position x in the image, as:

$$z(x) = y(x) + \sigma(y(x))\xi(x) \qquad x \in X \tag{1.7}$$

where $\xi(x) : X \to \mathbb{R}$ is zero-mean random noise with unitary variance and the function $\sigma : \mathbb{R} \to \mathbb{R}^+$ gives the standard deviation of the total noise component.

The noise model in Equation (1.7) can be expressed with two mutually independent components [57]:

$$\sigma(y(x))\xi(x) = \eta_p(y(x)) + \eta_g(x) \qquad x \in X \tag{1.8}$$

where η_p is the Poissonian part, and η_g is the Gaussian part, characterized as:

$$\chi(y(x)+\eta_p(y(x))) \sim \mathcal{P}(\chi \cdot y(x)) \qquad \chi > 0,\ x \in X$$
$$\eta_g \sim \mathcal{N}(0, b) \qquad b \geqslant 0,\ x \in X$$

The variance of η_p is proportional to the value of the original image $y(x)$; the Gaussian component η_g has a constant variance equal to b which, along with χ, depends on the hardware characteristics of the sensors.

1.2 Image Denoising

The image denoising problem can be formulated as estimating the noise-free image y from its observed noisy image z, which is corrupted by additive noise η (as in Eq. (1.2)). The main challenge facing any denoising algorithm is to suppress noise artifacts while retaining finer characteristics, details, and edges, in the image. In the algorithms described throughout this thesis, we assume the AWGN model.

Image denoising is an additive decomposition problem: the task is to decompose a noisy image into a denoised image component and a noise component, and we are interested in finding a *plausible* denoised image [13]. The idea is that a denoised image should resemble the original image, and a noisy component should agree with the noise model. The proposition of plausibility therefore involves *prior knowledge*, i.e. one has information about the image (e.g. regularity, smoothness, etc) and the noise (e.g. statistical distribution, etc).

1.3 Evaluation of Denoising Results: Quality Measures

Image quality metrics can be divided into three broad categories: (i) full-reference, (ii) no-reference, and (iii) reduced-reference metrics. The full-reference metrics require that the true underlying image is available in order to compute an evaluation measure, whereas no-reference metrics perform a "blind" quality assessment, i.e. the true underlying image is not available. Reduced-reference metrics presumes that the true image is partially known.

The objective of any denoising algorithm is to provide an estimate $\hat{y}$ of the original image y from the noisy observation z. After denoising an image, the performance of the algorithm needs to be quantified. For this, the full-reference quality metrices described in this thesis are Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity (SSIM) index.

1.3.1 Mean Squared Error

The MSE corresponds to the expected value of the squared error loss. The error is the magnitude of the dissimilarity between the original signal and the estimated one. In general, when y is an image, defined for all $x \in X \subset \mathbb{Z}^2$, and $\hat{y}$ is its estimate, this measure is defined as:

$$\mathrm{MSE} = \mathbb{E}[(y - \hat{y})^2] \approx \frac{1}{|X|} \sum_{x \in X} (y(x) - \hat{y}(x))^2 \tag{1.9}$$

where $|X|$ is the total number of pixels. It is one of the most commonly used metric for image quality assessment.

1.3.2 Peak Signal-to-Noise Ratio

The PSNR is the ratio between the maximum power of a signal and the power of the corrupting noise. PSNR, usually measured on a logarithmic scale (dB) scale due to a signal's wide dynamic range, is related to the MSE as follows:

$$\mathrm{PSNR} = 10 \log_{10} \frac{M^2}{\mathrm{MSE}} \tag{1.10}$$

where M is the maximum possible value of the signal, i.e. for images with an intensity range of [0-255], $M = 255$. An improvement of 1 dB corresponds to approximately 20% reduction in MSE.

PSNR and MSE are useful fidelity measures, but do not always serve as a good indicator of the visual quality of the estimated image.

1.3.3 Structural Similarity Index

The SSIM index exploits known characteristics of the human visual system. SSIM is a full-reference image quality metric which separates the task of similarity measurement into three components: (i) luminance, (ii) contrast, and (iii) structure:

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)\,(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)\,(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{1.11}$$

where σ is the standard deviation, and μ is the mean intensity value, of the image; C is a constant used to avoid instability when the denominator is close to zero.

SSIM takes into account the fact that the human visual system is sensitive to relative changes in luminance, rather than to absolute changes in luminance. It also considers image degradation as a perceived change in structural information, where structural information is the idea that spatially close pixels have strong inter-dependencies and are correlated [69].

For the purpose of this thesis, both PSNR and SSIM will be used as the reference quality measures.

2. SELF-SIMILARITY IN NON-LOCAL IMAGE DENOISING

Non-local self-similarity is widely acknowledged as an effective regularization *prior* for natural images. The utilization of non-local self-similarity in image processing gained prominence with the fractal model of coding for natural images [45], where it was demonstrated that natural images could be compressed by expressing their self-similarity as "self-transformability on a block-wise basis". In [73], the authors developed a novel approach to image filtering by exploiting the long-range correlation in natural images. Currently, non-local denoising and, more specifically, patch-based algorithms have become an established paradigm; and have been successfully applied to a wide range of imaging problems.

Patch-based Self-Similarity: For the purpose of explanation, we refer the reader back to the observation model stated in Eq. (1.2). Also, for the sake of simplicity, we assume that the images can be extended beyond the boundary of X to the whole $\mathbb{Z}^2$ through any standard padding technique.

Let $U \subset \mathbb{Z}^2$ be a neighbourhood centered at the origin, then the patch centered at a pixel $x \in X$ in the noisy observation z can be defined as:

$$\mathbf{z}_x(u) = z(u + x) \qquad u \in U \tag{2.1}$$

Similarly, we define the noise-free patches as:

$$\mathbf{y}_x(u) = y(u + x) \qquad u \in U \tag{2.2}$$

Natural images are, generally, highly redundant and a non-local algorithm utilizes these similarities to estimate the expected value of an image patch. By this, we mean that every patch in a natural image has a large number of mutually similar patches, located at different spatial positions, as shown in Figure 2.1. The Euclidean distance between the pixel intensities is used to assess the patch similarity, and is therefore dependent on the patch size. Large patches are encouraged for their robustness to noise, but using a larger patch will hinder the algorithm from finding redundancies, especially if the

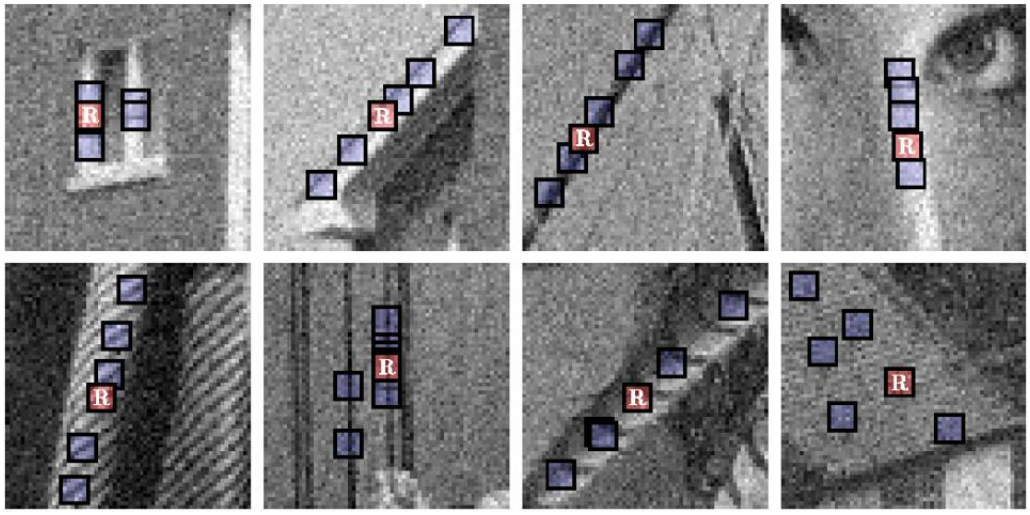


Figure 2.1: Self-similarity in natural images: for a given reference patch **R** , there exists many similar patches at different spatial locations (Reproduced from [22]).

image has textures with distinctive transitions, or curved and contrasted edges [30]. Thus, to compensate between these two opposing ideas, a windowed Euclidean distance is very often used. This distance has an inverse relation to the similarity, i.e. patches with a larger distance contribute less to the final estimate of reference patch and vice-versa. Hence, the efficacy of an algorithm is dependent on the validity of the underlying metric model.

2.1 Non-local Means Denoising

The NL-means algorithm, introduced in [6], is a non-local filter that aims at removing noise, without undermining the useful information in the original image, by exploiting redundancy and self-similarity inherent in a natural image. The general concept of non-local means is to estimate a "reference" pixel in the noisy image z as the weighted average of all pixels whose neighbourhood is similar to the neighbourhood of the reference pixel. The weights are calculated as a function of similarity between the neighbourhood of a reference pixel and the neighbourhood associated with every other pixel in the image [7]. The difference between this method and other adaptive spatial domain filtering methods is that this algorithm does not presuppose a locality constraint. In its basic implementation, the NL-means follows the given formulation:

$$\hat{y}(x_1) = \sum_{x_2 \in X} w(x_1, x_2) z(x_2) \qquad \forall x_1 \in X \tag{2.3}$$

where $\{w(x_1, x_2)_{x_2 \in X}\}$ is the set of adaptive weights that depend on the similarity between the image intensities of pixels x_1 and x_2, as detailed further below.

2.1.1 Weight Function

The weight function in (2.3) is normalized as follows:

$$0 \leqslant w(x_1, x_2) \leqslant 1 \tag{2.4}$$

$$\sum_{x_2 \in X} w(x_1, x_2) = 1 \tag{2.5}$$

To further understand the concept of similarity between a pair of pixels (x_1, x_2) in a given noisy image z, for $X \subset \mathbb{Z}^2$, we define the idea of a neighbourhood on X , which can have varying shapes and sizes to better adapt to the image.

Definition 2.1. *A neighbourhood on X is a family $\mathcal{N} = \{\mathcal{N}_x\}_{x \in X}$ of subsets of X such that $\forall x \in X$ the following conditions hold:*

1. $x \in \mathcal{N}_x$ *; and*
2. $x_0 \in \mathcal{N}_{x_1} \Rightarrow x_1 \in \mathcal{N}_{x_0}$

The set $\mathcal{N}_x \subset X$ is called the neighbourhood (nbd.) of x.

The limitation of z to a neighbourhood $\mathcal{N}_x$, denoted by $z(\mathcal{N}_x)$, is:

$$z(\mathcal{N}_x) = \{z(x), \quad x \in \mathcal{N}_x\} \tag{2.6}$$

where $z(\mathcal{N}_x)$ is a vector of pixels and $\mathcal{N}_x$ defines the neighbourhood of pixel x, which is normally a square-block of pre-defined size. The similarity between two pixels (x_1, x_2) is a function of the similarity of the intensity gray level vectors $\mathcal{N}_{x_1}$ and $\mathcal{N}_{x_2}$. The pixels with a similar gray level neighbourhood to $\mathcal{N}_{x_1}$ will have larger weights assigned to them.

One possible solution to the problem of computing the similarity of two pixels (and by extension, patches) is the Gaussian weighted Euclidean distance. This consists of taking the sum of squared differences between the two patches, weighted with a Gaussian kernel $\mathcal{G}_\alpha$ having a pre-defined standard deviation α:

$$d(x_1, x_2) = \|z(\mathcal{N}_{x_1}) - z(\mathcal{N}_{x_2})\|^2_{2,\mathcal{G}_\alpha} = (\mathcal{G}_\alpha * |z(\mathcal{N}_{x_1}) - z(\mathcal{N}_{x_2})|^2)(0) \tag{2.7}$$

The distance operator is defined as the *windowed quadratic distance* between image patches centered at x_1 and x_2, respectively. It was shown in [27], that the L^2 distance is a reliable measure for the comparision of image patches in a texture window. This measure is also more adapted to the white Gaussian noise in z:

$$\mathbb{E}\left[\|z(\mathcal{N}_{x_1}) - z(\mathcal{N}_{x_2})\|^2_{2,\,\mathcal{G}_\alpha}\right] = \|y(\mathcal{N}_{x_1}) - y(\mathcal{N}_{x_2})\|^2_{2,\,\mathcal{G}_\alpha} + 2\sigma^2 \tag{2.8}$$

where σ^2 is the variance of the noise η corrupting the original signal y. This equality shows that, in expectation, the Euclidean distance preserves the order of similarity between pixels. So the most similar pixels to x in z are also expected to be the most similar pixels of x in y, as shown in Fig 2.2.

Gray-scale Image (Reproduced from [6]). Color Image (Reproduced from [9]).

Figure 2.2: Illustration of weight function based on similar patches.

We can now formally define the weight function, $w(\,\cdot\,,\,\cdot\,)$ as:

$$w(x_1,\,x_2) = \frac{1}{C(x_1)} \exp \frac{-\|z(\mathcal{N}_{x_1}) - z(\mathcal{N}_{x_2})\|^2_{2,\,\mathcal{G}_\alpha}}{h^2} \tag{2.9}$$

where $h > 0$ is a filtering parameter controlling the decay of the exponential function in the weights, $\|\cdot\|$ is the Gaussian weighted distance, and the term $C(x_1)$ is a normalizing factor which guarantees the weights w will satisfy the conditions given in Equations (2.4) and (2.5), i.e.

$$C(x_1) = \sum_{x_2 \in X} \exp \frac{-\|z(\mathcal{N}_{x_1}) - z(\mathcal{N}_{x_2})\|^2_{2,\,\mathcal{G}_\alpha}}{h^2} \tag{2.10}$$

The procedure assigns larger weights to the terms $z(\,\cdot\,)$ in Eq. (2.3) that correspond to pixels belonging to similar patches (i.e. where the pixel intensity difference between patches $d(x_1,\,x_2)$ is small), regardless of their location within the image. The similarity between pixel intensities is estimated as a decreasing function of the Euclidean distance between patches. Hence, large Euclidean distances lead to small weights and vice-versa.

2.1.2 Filtering Parameter

The filtering parameter h, which controls the amount of blurring introduced in the denoising process, has been subject to intense scrutiny ever since its inception in [6]. The authors had suggested that the parameter h could be selected as the standard deviation σ of the noise in the image, a known priori. The experimental results obtained had reasonably good visual quality.

A number of authors [54] use a χ^2 test to set the parameter h. This leads to a linear relation between h and σ, and the experiments reported in [67] confirm that in terms of the PSNR, the best value of h is roughly proportional to σ. The value of the filtering parameter writes $h = k\sigma$, and the visual difference between the results with optimal h and the predicted value $k\sigma$ is not significant.

As has been shown in [26], if h is too small, the noise removal may not be effective. Conversely, if h is too large, the image will be over-smoothed.

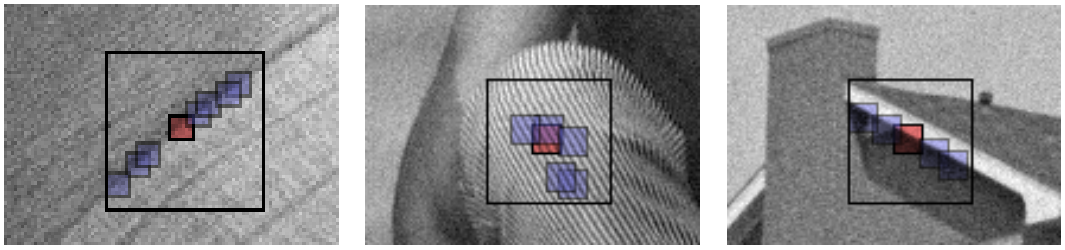


Figure 2.3: Illustration of concept of search neighbourhood, $\mathcal{N}_x$ (Reprint from [13]).

In practice, the averaging process is not performed over the entire image but instead, a "search neighbourhood" centered at x is used. Using a small search neighbourhood is common practice not only for computational reasons but, as is shown in [33], the denoising performance decreases as the search neighbourhood increases beyond a certain size.

A large patch allows a more robust discrimination between noisy areas which are not actually similar. The best visual and theoretic results, for high noise levels, are obtained with a large patch size [26].

2.1.3 NL-means Denoising for Color Images

The extension of NL-means algorithm from grayscale to color images is very straightforward, with a few minute differences. The windowed quadratic distance - previously defined in Eq. (2.7) - makes the weight distribution adapt to the local geometry of the image, as detailed in [11]. The NL-means algorithm is applied to color images by replacing the absolute value of the pixel intensity difference with the norm of the color difference vector:

Noisy Image

Denoised Image

Noisy Image

Denoised Image

Figure 2.4: Application of NL-means to image corrupted with Gaussian noise with standard deviation = 25/255.

$$d(x_1, x_2) = (\mathcal{G}_\alpha * \|z(\mathcal{N}_{x_1}) - z(\mathcal{N}_{x_2})\|^2)(0) \tag{2.11}$$

The averaging configuration given in Eq. (2.3), with the updated d-distance, is applied separately to the three color channels. So for each pixel, each channel value is the result of the weighted average of pixels having similar intensities. Compared to the grayscale case, the denoising results improve dramatically on color images because similar pixels are more effectively identified with three components.

The NL-means algorithm represented a paradigm shift in image denoising and inspired several powerful algorithms in the following years, such as BM3D and SAFIR [49]. For a comprehensive overview, we refer the reader to article [48].

2.2 BM3D

Block-matching and 3D-filtering (BM3D) is widely considered the state-of-the-art algorithm in terms of PSNR and subjective quality for images corrupted by white Gaussian noise. The algorithm utilizes the notion that natural images consist of self-similar patches. These similar two-dimensional patches [1] in the image are grouped, using a method called block-matching. The resulting stack of patches is a three-dimensional array, referred to as *groups*. Each group is processed by applying a linearly separable 3-D transform to obtain a sparse representation of the image, i.e. one that can be entirely described using a small set of coefficients; the image is disassembled with respect to elementary *basis* functions. The resulting coefficients are "shrunk", by thresholding the coefficients of the transformed domain, followed by an inverse of the 3-D transform. This strategy has been experimentally shown to be an effective way of detecting textures, edges, etc., in images, without losing much of the distinctive attributes [20]. Also, we note that an image patch can belong to several groups, which is different from clustering, where each patch can belong to only one cluster.

We assume the observation model described in Eq. (1.2). The method exploits both the spatial and frequency information of an image. The BM3D algorithm:

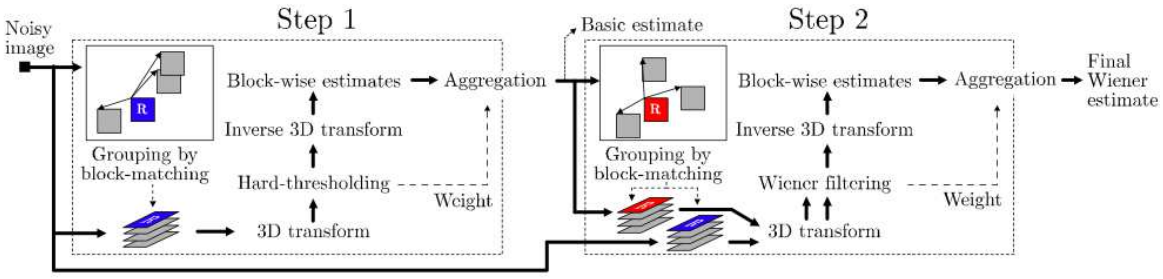


Figure 2.5: BM3D Algorithm Flowchart (Reproduced from [22]).

2.2.1 Algorithm

Let Z denote a square patch of fixed size $N \times N$ extracted from the noisy observation z. The group of patches is denoted by $\mathbf{Z}_S$, where $S \subseteq X$ is the set of coordinates identifying the patches Z_x grouped in $\mathbf{Z}_S$:

$$\mathbf{Z}_S = \{Z_x : x \in S \subseteq X\} \tag{2.12}$$

where x is the pixel coordinate of the top-left corner of the patch.

[1] Existing literature refers to these patches as either *fragments*, *neighbourhoods* or *blocks*, depending on the number of dimensions.

The global flowchart of the BM3D algorithm is shown in Fig. 2.5. A general outline of the BM3D algorithm is [19]:

i. For a given reference patch Z_R, find all similar candidate patches Z and stack them in a three-dimensional array $\mathbf{Z}_S$, which is the group; then,

ii. Perform collaborative filtering on the transformed group, and insert back the denoised two-dimensional estimates of all the grouped patches to the location where their noisy counterparts were found;

iii. Finally, for a given pixel x, there exists multiple estimates of the overlapping patches, which are then aggregated to produce the final image $\hat{y}(x)$

The BM3D algorithm repeats the above procedure in two different steps. The steps differ in the implementation of the collaborative filtering, in how similar-looking patches are found and how the coefficients are shrunk. The first step uses a collaborative hard-thresholding to produce a basic estimate of the original image y. The initial denoised image is then used as a reliable guide, or pilot, estimate of the ground-truth for a Wiener filtering operation.

In the first step, similar-looking patches are found in the noisy image itself, via block-matching. Robustness of block matching is improved by applying a normalized two-dimensional linear transform. The patches are then pre-filtered (hard-thresholded) in order to diminish the effect of noise. Formally, the distance between the reference patch Z_R and the candidate patch Z is defined as:

$$d(Z_R, Z) = \frac{\|\Gamma(\mathcal{T}_{2D}(Z_R)) - \Gamma(\mathcal{T}_{2D}(Z))\|_2^2}{N^2} \tag{2.13}$$

where Γ is a hard-threshold operator, N^2 is the number of pixels in the image patch, $\mathcal{T}$ is a orthonormal two-dimensional linear transform. The orthogonality ensures that the distance coincides with the ℓ^2 -distance of the denoised patch estimates in the space domain. Thus, the similarity can be computed directly from the spectral coefficients, without applying an inverse transformation.

The result of block-matching is a three-dimensional array constructed by grouping the reference patch Z_R with all the candidate patches Z, for which $d(Z_R, Z)$ is smaller than a predefined threshold. To reduce the computationally intensive nature of block-matching, the authors use a few practical tricks [23]: (i) Search for candidate patches is restricted to a search neighbourhood instead of the whole image, (ii) Not every image patch is used as a reference patch: the method skips a few pixels between successive reference patches. In practice, the algorithm is fast.

During the collaborative filtering step, a 3-D transform is applied to the obtained groups $\mathbf{Z}_S$. The 3-D linear transforms is formed by a separable composition

of a 2-D linear transform with a 1-D transform. The 2-D transform, e.g. DCT-transform or biorthogonal wavelet transform, has the effect of exploiting intra-patch correlations. The 1-D transform, e.g. Haar-transform, applied along the stacking dimension of the block, has the effect of exploiting inter-patch correlations. Thus, the image information will be concentrated into few 3-D transform coefficients. This is followed by the shrinkage of coefficients. It should be noted that the noise is white both before and after application of the transforms, given that the grouped blocks are non-overlapping and the transforms are orthonormal:

$$\hat{\mathbf{Y}}_S = \mathcal{T}_{3D}^{-1}\left(\Upsilon\left(\mathcal{T}_{3D}\left(\mathbf{Z}_S\right)\right)\right) \tag{2.14}$$

where Υ is a hard-threshold operator dependent on σ, $\hat{\mathbf{Y}}_{S_x}$ for $x \in X$ is the set of filtered groups, and $\mathcal{T}_{3D}$ is the normalized 3-D transform being adopted. And:

$$\hat{\mathbf{Y}}_S = \left\{\hat{Y}_x : x \in X\right\} \tag{2.15}$$

The denoised patches have to be re-inserted to their original places in the image. However, we have an over-complete representation of the initial denoised image, i.e. we have overlapping patch-wise estimates, along with several groups containing an estimate of the same patch. Thus, the final estimate of the pixel is obtained by aggregating the individual denoised patches. This is performed by a weighted averaging of the individual results, where the weight assigned is inversely proportional to the total sample variance of the corresponding patch-wise estimate. Now, patches from the denoised estimate are utilized in the next step.

In the second step, we exploit the denoised image *basic estimate* $\hat{y}^{basic}$ obtained during the first step. Using block-matching, two groups are formed for both the original noisy image z and the corresponding denoised estimate $\hat{y}^{basic}$. Assuming the noise in $\hat{y}^{basic}$ is relatively small, the thresholding operator Γ is removed to simplify the distance metric in Equation (2.13). The updated metric also replaces the d-distance with a normalized Euclidean ℓ^2-distance between patches extracted from the basic estimate:

$$d(Z_R, Z) = \frac{\|\hat{Y}_R^{basic} - \hat{Y}^{basic}\|_2^2}{N^2} \tag{2.16}$$

Hence, groups are formed of the basic estimate $\hat{\mathbf{Y}}_b$, and of the noisy image Z.

The same 3-D transform is applied to both the set of groups, $\hat{\mathbf{Y}}_b$ and Z; it should be noted that the 2-D and 1-D transforms need not necessarily be the same for both the steps, even though the qualitative difference for various choices is relatively minor. Wiener filtering is then used to achieve shrinkage, computed from the three-dimensional spectrum of the basic estimate group:

Noisy Image | Denoised Image

Noisy Image | Denoised Image

Figure 2.6: Application of BM3D to image corrupted with Gaussian noise having standard deviation = 25/255.

$$\mathbf{W} = \frac{|\mathcal{T}_{3D}(\hat{\mathbf{Y}}_b)|^2}{|\mathcal{T}_{3D}(\hat{\mathbf{Y}}_b)|^2 + \sigma^2} \tag{2.17}$$

where $\mathcal{T}_{3D}$ is the 3-D transform, σ^2 is the variance of the noisy image, and $\mathbf{W}$ are the Wiener shrinkage coefficients. Wiener filtering and inverse 3-D transform are then applied on the noisy group Z:

$$\hat{\mathbf{Y}} = \mathcal{T}_{3D}^{-1}\left(\mathbf{W} \cdot \mathcal{T}_{3D}\left(Z\right)\right) \tag{2.18}$$

where $\hat{\mathbf{Y}}$ is the final estimate of the set of patches. Finally, the individual patch estimates are aggregated by a weighted averaging to produce the final denoised image $\hat{y}$. Despite numerically superior results, the method is not yet perfect [16].

2.3 DDID

Dual-Domain Image Denoising (DDID) is a hybrid denoising algorithm implemented in both the spatial and transform domains, whose image quality rivals that of BM3D. The algorithm combines two classical filters from both the domains, and applies it over the image in a sequential manner. In the spatial domain, the bilateral filter is used; and for the transform domain, the short-time Fourier transform (STFT) is used. The bilateral filter preserves edges and other high-contrast features, whereas the STFT preserves details and textures [50].

Assuming the observation model described in Equation (1.2), the noisy image z is separated into two images - a high-contrast image and a low-contrast image - and then denoised separately. The high-contrast image is obtained by denoising the noisy image using the bilateral filter. The residual of the bilateral filtering is the low-contrast image [29], which is denoised using coefficient shrinkage in the transform domain. Thus, the original image can be approximated as the sum of the two denoised images:

$$\hat{y} = \hat{s} + \hat{S} \tag{2.19}$$

where $\hat{s}$ and $\hat{S}$ are the denoised high- and low- contrast images, respectively. The algorithm depends on iterative learning, i.e. the denoised result of an iteration is used as a *guide* for the subsequent iteration.

2.3.1 Spatial Domain: Bilateral Filter

First, a short account of the Bilateral filter before proceeding to its implementation.

Bilateral Filter: The filter attempts to smooth an image while preserving edges. This is achieved by choosing pixels based not only on their spatial proximity, but also on their intensity similarity [65]. Therefore, both the spatial distance and the intensity distance are important for determining the weights.

Given the observation model stated in Eq. (1.2), the output of the bilateral filter can be formulated as in Eq. (2.3). In the given noisy image z, for pixel locations x_1 and $x_2 \in \mathcal{N}_{x_1}$, the weight function is defined as:

$$w(x_1, x_2) = \frac{1}{C(x_1)} \exp\left(\frac{-\|x_1 - x_2\|^2}{2\sigma_s^2}\right) \exp\left(\frac{-|z(x_1) - z(x_2)|^2}{\gamma_r\, \sigma^2}\right) \tag{2.20}$$

where σ_s and γ_r are parameters controlling the decay of weights in spatial and intensity domains, respectively. For a spatial patch $\mathcal{N}_{x_1}$ of pixel x_1, we define the normalizing factor as:

$$C(x_1) = \sum_{x_2 \in \mathcal{N}(x_1)} \exp\left(\frac{-\|x_1 - x_2\|^2}{2\sigma_s^2}\right) \exp\left(\frac{-|z(x_1) - z(x_2)|^2}{\gamma_r\, \sigma^2}\right) \tag{2.21}$$

The optimal value of the hyper-parameters σ_s and γ_r are image-dependent, and also depend on the level of noise σ . Furthermore, the size of the patch is selected by trial and experiments. The bilateral filter is widely used in imaging applications due to the simplicity of its concept.

Now, in DDID, instead of having just the noisy input image, there is an additional guide image g that defines the bilateral filter used in the algorithm. The authors define the bilateral kernel using g by measuring patch (or, structure) similarity. The joint bilateral filter [40] is applied simultaneously on both the guide g and noisy image z to obtain the denoised high-contrast images $\hat{g}(x)$ and $\hat{s}(x)$, respectively. The denoised high-contrast values at a pixel x is calculated as:

$$\hat{g}(x) = \frac{\sum\limits_{x_1 \in \mathcal{N}_x} k(x,\, x_1)\, g(x_1)}{\sum\limits_{x_1 \in \mathcal{N}_x} k(x,\, x_1)} \tag{2.22}$$

$$\hat{s}(x) = \frac{\sum\limits_{x_1 \in \mathcal{N}_x} k(x,\, x_1)\, z(x_1)}{\sum\limits_{x_1 \in \mathcal{N}_x} k(x,\, x_1)} \tag{2.23}$$

where the bilateral kernel, defined over a square patch $\mathcal{N}_x$ with length r , is:

$$k(x,\, x_1) = \exp\left(\frac{-\|x - x_1\|^2}{2\sigma_s^2}\right) \exp\left(\frac{-|g(x) - g(x_1)|^2}{\gamma_r\, \sigma^2}\right) \tag{2.24}$$

The spatial kernel, whose shape is decided by σ_s , removes periodic discontinuities. The range kernel, whose shape is decided by γ_r , finds similar patches.

2.3.2 Transform Domain

The difference of the bilaterally filtered high-contrast values from the current (iteration) guide and noisy values, i.e. $\hat{g}(x)$ from $g(x)$ and $\hat{s}(x)$ from $z(x)$ respectively, gives us the low-contrast signals. The extracted signals are blurred with the noise-free range kernel, given in Eq. (2.24), to smooth any fluctuation in intensities. The STFT [3] of both the signals is performed by combining the spatial Gaussian kernel from the bilateral filter together with the DFT, to give the frequency domain coefficients $G(x, f)$ and $S(x, f)$. These resulting coefficients are defined for the frequencies f over the frequency window $\mathcal{F}_x$ having size $\mathcal{N}_x$. The value of the noisy Fourier coefficients are given as:

$$G(x,f) = \sum_{x_1 \in \mathcal{N}_x} \exp\left(\frac{-i2\pi(x_1 - x)\cdot f}{(2r+1)}\right) k(x,\, x_1)\,(g(x_1) - \hat{g}(x)) \tag{2.25}$$

$$S(x,f) = \sum_{x_1 \in \mathcal{N}_x} \exp\left(\frac{-i2\pi(x_1 - x)\cdot f}{(2r+1)}\right) k(x,\, x_1)\,(z(x_1) - \hat{s}(x)) \tag{2.26}$$

The Fourier transform dictates that the noise in every pixel of the image is evenly distributed over all frequencies. Thus, every frequency of S has Gaussian noise with the same variance:

$$\sigma^2_{x,f} = \sigma^2 \sum_{x_1 \in \mathcal{N}_x} k^2(x,\, x_1) \tag{2.27}$$

The STFT recovers previously lost detail features, and is unaffected by edges.

2.3.3 Frequency Domain: Coefficient Shrinkage

The noisy Fourier coefficients $S(x,f)$ are denoised by using shrinkage factors which are inversely proportional to the range kernel ((2.24)) used in the bilateral filter. The rationale behind choosing an inverse relation is to ensure the coefficient shrinkage factor $K(x,f)$ retains the signal and discards the zero-mean noise. The discontinuities in the image were removed in the previous steps, and hence denoising in the Fourier domain does not introduce ringing artifacts. Then, the inverse DFT over the frequency domain $\mathcal{F}_x$ yields the denoised low-contrast value of the central pixel [72]. This value is the mean of all the shrunk coefficients:

$$\hat{S}(x) = \frac{1}{|\mathcal{F}_x|} \sum_{f \in \mathcal{F}_x} K(x,f)\, S(x,f) \tag{2.28}$$

where the coefficient shrinkage factors are defined using the spectral guide $G(x,f)$:

$$K(x,f) = \exp\left(-\frac{\gamma_f\, \sigma^2_{x,f}}{|G(x,f)|^2}\right) \tag{2.29}$$

The coefficient shrinkage parameter γ_f imitates the bilateral range parameter γ_r [50]. The inverse transform to recover the denoised value is repeated for every pixel of the image.

Noisy Image

Denoised Image

Noisy Image

Denoised Image

Figure 2.7: Application of DDID to image corrupted with Gaussian noise with standard deviation = 25/255.

3. FEATURES OF HUMAN VISUAL SYSTEM

The Human Visual System (HVS) consists of three functional organs, i.e. the eyes, optical nerves, and the brain. The eye can be considered to be the biological equivalent of a camera, as both are used to focus the incident light rays and for exposure control; the optic nerves send an electric signal, which is representative of the light rays, from the eye to the brain; and the brain is responsible for complex image processing tasks.

Light from external objects in the visual field is focused onto a light-sensitive screen, called the retina. An inverted 2-D retinal image is then transformed into the perceived 3-D image by the visual system. Visual perception is provided by several optical and neural transformations, for which we refer the reader to [68]. Visual perception is a subject of anatomy whereas visual cognition is studied in psychology. We present a short description of the main features of human vision.

3.1 Physiology of Vision

The retina is a light-sensitive detector at the inner surface of the posterior of the eye, and contains photoreceptor cells which absorb light rays. Photoreceptors convert light energy into electrochemical signals and can be functionally classified into *rods* and *cones*, named for their apparent shape. Rods are stimulated by light of lower intensity, whereas cones are stimulated by any one of a set of three different wavelengths of light. These wavelengths are characterized by red, green, and blue light, and correspond to specific bands in the electromagnetic spectrum. Hence, cones are fundamental to encoding color information and have high spatial acuity, whereas rods do not mediate color vision and have low spatial acuity [15].

The retina contains approximately 120 million rods and 7 million cones [42], and transmits electrochemical signals to the brain through the optic nerves. A detailed explanation of the different types of retinal cells can be found in [47] but we would like to mention *ganglion cells* which receive the output of the retina as signals, illustrated as *spikes*, and transmits them to the brain. The retinal cells are organized in layers, with connection bundles between each layer. These connectional layers are called electrical synapses, which are ionic channels allowing the bidirectional circulation of ions between the cells. These synapses result in a local sharing of the information between cells of the same type.

3.2 Central vs. Peripheral Vision

The retina exhibits a radial orientation bias, with the central region displaying a much higher density of cones and, therefore, a much higher spatial resolution compared to the periphery [64]. This central region is called the *fovea* and covers about 6° of the visual field. Our visual acuity is spatially non-uniform and is stronger for radial lines, i.e. along directions towards the center of the gaze.

The fovea, which is filled with color-sensitive cones, is at the core of mediating our acute vision. Conversely, rods are dispersed in the periphery of the retina. Fig. 3.1 shows the densities of both types of light receptors in a human retina, as a function of eccentricity r (distance from the center of retina). As observed, the density of cones follows a power law proportional to r^{-1} [37], and leads to our loss of spatial acuity at the retinal periphery.

The density of rods reaches its maxima at approx. 20° eccentricity, and then gradually decreases. The large number of rods allows the retina to detect subtle changes in illumination and movement. The non-uniform distribution of cells in the retina can be interpreted as an optical low-pass filter, thereby explaining the lack of details and color information as one moves away from the fovea.

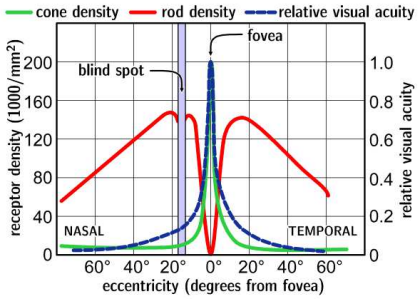


Photo-receptor density.

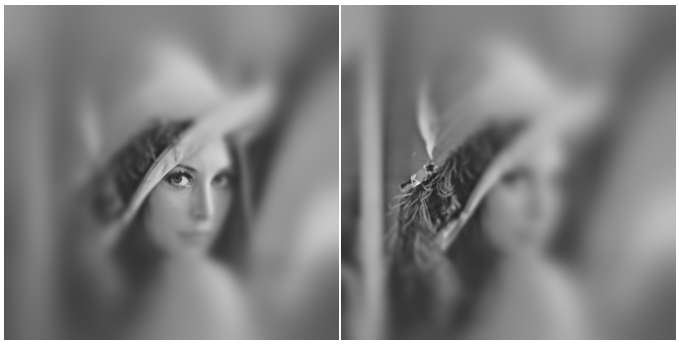


Figure 3.1: Left: Diagram for relative visual acuity vs. eccentricity [18]. Right: Examples of the *Lena* image foveated at two different fixation points [32].

3.2.1 Temporal Vision

Human visual response to motion is characterized by two distinct facts: *persistence of vision* and *phi-phenomenon*. The former describes the temporal representation of successively different images in the HVS, and the latter creates the illusion of motion between those images [41]. Both these features are utilized in television, theaters, etc., as the brain conceptually completes the gaps between frames.

Persistence of vision also describes the inability of the retina to sample rapidly changing intensities. A stimuli flashing at approx. 24-48 frames per second appears steady, depending on luminance and contrast conditions [24]. This is known as

Critical Fusion Frequency (CFF) and explains why the eye is more sensitive to flicker at higher luminance than lower luminance.

Experiments show that the temporal response of the HVS to motion is not consistent across the visual field. The central field of view (fovea) is more receptive to slower motion whereas the periphery is more sensitive to faster motion, although motion is discerned uniformly across the visual field. Given the salient nature of movements at the periphery, its main task is motion detection.

3.3 Receptive Field

The receptive field of a retinal cell is the area of the visual field where stimulation by packets of photon, or light, leads to the firing of the cell. Given a retinal cell centered at $(x_0,\ y_0)$ on the retina, its receptive field **RF** can be thought of as a convolution kernel satisfying the following mathematical framework:

$$A(t) = \int_{u=0}^{+\infty} \int_{(x,y)\in\mathbf{RF}} I(x_0 - x,\ y_0 - y,\ t - u)\, K(x,\ y,\ u)\, dx\, dy\, du \tag{3.1}$$

where $I(x,\ y,\ t)$ is the luminance profile of the image, $K(x,\ y,\ t)$ is the linear receptive field of the cell, and $A(t)$ is a measure of activity appropriate to the type of cell. Receptive fields can detect contrast changes within an image, revealing edges or shadows. In fact, the operation performed in Eq. (3.1) corresponds to:

$$A(t) = (I * K)(x_0,\ y_0,\ t) \tag{3.2}$$

where $*$ denotes spatio-temporal convolution. Since convolution corresponds to multiplication in the Fourier space, the Fourier analysis is a suitable approach for the study of such a filter.

3.3.1 Separability of Space and Time

To separately study the spatial and temporal properties of the cell, we consider an input stream of images each defined at time t. The temporal behaviour of the cell, in response to an image given by $I(t)$, can be linearly studied as:

$$A(t) = \int_{u=0}^{+\infty} I(t - u)\, K_{\text{temporal}}(u)\, du \tag{3.3}$$

Similarly, the spatial behaviour of the cell's activity, in response to a static image $I(x,\ y)$, can be linearly studied as:

$$A = \int_{(x,y)\in\mathbf{RF}} I(x_0 - x,\ y_0 - y)\, K_{\text{spatial}}(x,\ y)\, dx\, dy \tag{3.4}$$

In practice, the receptive field includes the type, size and shape of stimulus needed to cause maximal response. This makes it problematic to express the best-fitting linear kernel $K(x,\ y,\ t)$ as a product $K_{\text{spatial}}(x,\ y)\ K_{\text{temporal}}(t)$, because of the influence of experimental conditions [70]. Hence, all reductions to study either temporal or spatial properties necessitate a loss of information.

3.3.2 Center-Surround Architecture

Receptive fields have a characteristic center-surround architecture which is known for retinal filtering and detecting strong spatial contrast, such as object edges [52]. When light impinges upon a spot in the ganglion cell's receptive field, it elicits different responses depending on the precise location of the spot:

- If the spot is at the *center* of the receptive field, it leads to an increased firing of spike signals by the ganglion cell.
- If the spot is at the *surround* of the receptive field, it has an inhibitory effect on the firing of spikes by the ganglion cell.

The center-surround architecture is not a static feature in our retinas. The activity pattern of the ganglion cells varies drastically with the size and influence of the center and surround receptive fields.

3.3.3 Difference of Gaussians

The center-surround architecture of receptive fields is approximated by a filter consisting of a Difference of Gaussians (DoG), as is shown in [62]:

$$K_{\text{spatial}}(x,\ y) = w_c G_{\sigma_c}(x,\ y) - w_s G_{\sigma_s}(x,\ y) \tag{3.5}$$

where w_c and w_s are the weights of the *center* and *surround* components of the receptive field, respectively, and $G_\sigma(x,\ y)$ is a normalized, two-dimensional Gaussian function with standard deviation σ:

$$G_\sigma(x,\ y) = \frac{\exp\left(\frac{-(x^2+y^2)}{2\sigma^2}\right)}{2\pi\sigma^2} \tag{3.6}$$

The Difference of Gaussians approximation of retinal receptive fields is illustrated in Fig.3.2a. It is believed that the human retina extracts details from images using DoG of various sizes and encodes such differences with an activity potential. Hence, the eye does not perceive the absolute luminance level, but only the relative luminance values.

3.3.4 Parameter Fitting for DoG Model

The DoG model in Section 3.3.3 is contingent on four parameters, with the following functional interpretation [28]:

Spatial Resolution: given by σ_c , it is the amount of blur applied to the image formed on the retina, and gives the cut-off frequency of the retinal filtering.

Linear Gain: given by w_c, it gives the order of magnitude for retinal amplification - from input luminance to spiking activity.

Relative *Surround* Extent: given by (σ_s / σ_c) , it expresses the approximation of the best-fitting Gaussian for the *surround* across the ganglion cells.

Relative *Surround* Weight: given by (w_s / w_c) , it helps biologically determine the best approximation for weights in *center* and *surround* areas.

3.4 Color Vision

Color is created by utilizing two properties of light, energy and wavelength (or, frequency of vibration). Color vision combines both the amount of energy and wavelength composition reflected from an object to detect it [71]. The wavelength contrast helps us to perceive the color of the object, whereas the energy content helps us perceive the luminance of the object. Minus the luminance, the objective quality of a color with respect to its wavelength is known as chromaticity.

Photoreceptors, i.e. rods and cones, are neurons specialized to detect light. Rods are very sensitive but their response saturates as the light levels increase. Cones are less sensitive but can adapt to the increasing light levels, and are almost impossible to saturate. The 3 types of cones discussed in Sec. 3.1, based on their dominant wavelength as either red (R), green (G), or blue(B), are the basis of trichromatic vision. This trichromacy of vision facilitates the perception of color by linearly combining the responses from the different cones and is primarily based on color mixing experiments.

Color vision is of two types - foveal and peripheral. There is an unanimous agreement in the research community about the functioning of the foveal color vision, whereas there is a lack of consensus and understanding about the functioning of the peripheral color vision [17], [60]. Exacerbating the situation is the dearth of experimental data, and absence of quantified performance parameters, for measuring peripheral color discrimination. Behavioural studies opine that the distribution of cones is responsible for foveal color vision, whereas rods cannot discriminate colors which is why we are more sensitive to shades of grey at the periphery [66]. A study to describe the peripheral color vision in terms foveal color vision via color matching was described in [59].

3.5 Foveated Imaging

The retinal image has many spatially variant characteristics, due to the arrangement of photo-receptive cells. The image is sharpest at the center of the gaze (fixation point, Fig. 3.1) and becomes progressively blurrier as the distance from the center increases. This phenomenon is termed as foveated vision, foveated imaging, or foveation. Foveation is the result of a cascade of certain space-variant optical, sampling, and processing contributors [46].

Light entering the eye is focused on the retina by the cornea. This biological optical system provides high acuity and accuracy in the fovea region and low acuity at the periphery of the visual field, i.e. perifoveal region. This can be achieved by imposing a spatially-variant blur on the input patch, such that full detail is kept in the central section (fovea), while the peripheral parts are defocused by means of a convolution with a set of low-pass Gaussian filters, to remove fine details [25]. Moreover, there are fewer ganglions as we move further away from the fovea, and each of them are connected to more photoreceptors, and this gradually decreases the level of spatial acuity.

Given the anatomical and psychological explanations for foveation, we adopt the simplest framework to implement the low-pass filter in our work. We conduct experiments to show the functional advantage of foveation in imaging algorithms.

3.6 FREAK

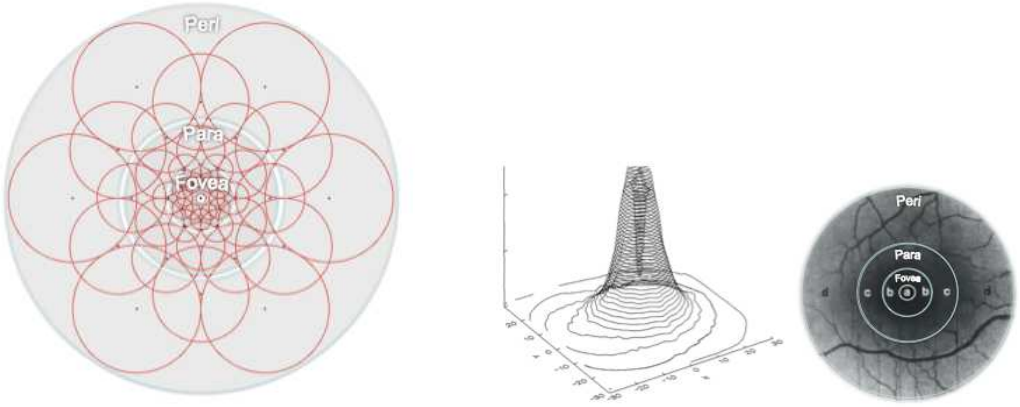
Fast Retina Keypoint (FREAK) is a robust image keypoint descriptor, designed according to the biological topology of ganglion cells in the eye. The image patches, utilized as the sparse and compact keypoint, are independent of noise, are scale and rotation invariant, and are employed in many computer vision and machine learning algorithms.

The FREAK descriptor encodes the responses of several pairs of receptive fields, which are obtained by the convolution of an image with Gaussian kernels having varying standard deviations. The descriptor then selects pairs based on an intensity similarity metric to decrease the degree of the descriptor. This process results in a structured pattern which resembles the short fluctuations in eye movements (or, saccades) of the HVS.

3.6.1 Retinal Sampling Pattern

Sampling grids are used to draw a comparision between pixel intensities in a pairwise manner. In [1], the authors use a (circular) retinal sampling grid having a high concentration of sample points in the central region, which gradually decreases as one moves towards the periphery of the grid, as shown in Fig. 3.2.

To reduce the dependence of the sample points on noise, they are Gaussian smoothed. The smoothing is done by applying Gaussian kernels of varying size, having an increasing blur as the distance from the keypoint increases, to the corresponding sample points. It was observed that varying the Gaussian kernels in accordance with the highly structured retinal pattern leads to an increase in the descriptors performance, which was further bolstered by utilizing the overlapping receptive fields.



(a) FREAK Sampling Pattern.

(b) Ganglion cell distribution in the retina.

Figure 3.2: Illustration of FREAK pattern similar to Ganglion cell distribution; where (a) is reproduced from [1] and (b) from [44]

Redundancy is added to use fewer receptive fields, and bring more discriminative power [14].

3.6.2 Coarse-to-fine Descriptor

The selected pairs of receptive fields are subtracted from their corresponding Gaussian kernel, and the results are thresholded to construct the binary descriptor F. The descriptor F is a string containing a sequence of one-bit binary-quantized Difference of Gaussians (DoG):

$$F = \sum_{0 \leqslant a < N} 2^a \, T(P_a) \tag{3.7}$$

where bit P_a is a pair of receptive fields, N is the desired size of the descriptor, and T performs the the pairwise intensity comparision tests as:

$$T(P_a) = \begin{cases} 1 & \text{if } (I(P_a^{r1}) - (P_a^{r2})) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.8}$$

with $I(P_a^r)$ being the smoothed intensity of the receptive field of the pair P_a .

To learn the best pairs from the data, the following greedy algorithm [63] is used:

i. Construct a matrix D of extracted keypoints, where each row corresponds to a keypoint and each column corresponds to a descriptor consisting of pairs in the retinal sampling pattern.

ii. Calculate the mean of each column. A mean of 0.5 leads to the highest variance, and is thus a strong discriminating feature, for a binary distribution.

iii. Order the columns with respect to variance. Then, keeping the best column, iteratively add the remaining columns which have a low correlation coefficient with the existing columns.

There is a coarse-to-fine ordering in the structure of the selected pairs. A symmetric scheme is captured due to the orientation of the pattern, where the selected pairs are grouped into clusters [1]. The first clusters that are selected mainly compare sampling points in the peripheral receptive fields of the pattern, whereas the last clusters compare points in the centered receptive fields, which is similar to the behaviour of the human eye [2].

Saccadic Search: Saccades are rapid eye movements that continually reposition our gaze, in order to form a complete and detailed image of the surrounding environment. Stable and persistent perception of visual space requires that features in the new retinal image are associated with corresponding features in the previous retinal image [51].

The FREAK descriptor starts by parsing the first 128 bits of the descriptor, which contains the coarse information. If the distance between the first set of bits and the next set is smaller than a pre-defined threshold, the comparision is continued with successive bits to analyze finer information. Hence, a series of comparisions are performed to mimic the saccades.

First, peripheral receptive fields are used to estimate the location of an object of interest. Then, the validation is performed with more densely distributed receptive fields in the fovea area. The feature selection is heuristic and matches the radial model of the human retina.

3.6.3 Orientation

Descriptors consisting of pairs having symmetric receptive fields with respect to the center are selected. The local gradients of the selected pairs are cumulatively summed to estimate the orientation of the keypoint [53]. If G is the set of all the pairs used to compute the local gradients, then:

$$O = \frac{1}{M} \sum_{P_o \in G} (I(P_o^{r_1}) - (P_o^{r_2})) \frac{P_o^{r_1} - P_o^{r_2}}{\|P_o^{r_1} - P_o^{r_2}\|} \tag{3.9}$$

where M is the number of pairs in G and $P_o^{r_i}$ is the two-dimensional vector of the spatial coordinates of the center of receptive field. The descriptor is normalized with respect to dominant orientation, though it should be noted that the retinal pattern has larger receptive fields in the peripheral area, leading to more error in the orientation estimation and a larger memory load.

The experiments show that FREAKs are faster to compute, with lower memory load, and more robust than existing methods (e.g. SIFT, SURF or BRISK), especially for algorithms on smart phones.

4. FOVEATED SELF-SIMILARITY IN IMAGE DENOISING

Non-local self-similarity is exploited to correctly identify similar patches in a natural image, on the basis of a suitable patch distance. Patch similarity is, typically, assessed through the windowed Euclidean distance of the pixel intensities. In this thesis, we substitute the windowed distance with a foveated distance, which employs the Euclidean distance between foveated patches. Such patches are blurred by point-spread functions (PSFs) having an increasing standard deviation (and thus, an increasing *blur*) as the spatial distance from the center of the patch grows. We design specific foveation operators, motivated by the human visual system (HVS), to blur a patch so as to measure the patch similarity. If we consider the patch center as the point of fixation, then the foveated distance mimics the inability of the HVS to discern details at the periphery.

Foveated Self-Similarity: The foveated distance installs a different form of self-similarity in the context of non-local image modeling, the foveated self-similarity [32]. The reader's attention is referred to the observation model stated in Eq. (1.2), along with the assumption presented in Eq. (2.1).

A generalized definition of distance d, stated in Equation (2.7), is given by the windowed quadratic distance between patches centered at x_1 and x_2:

$$d(x_1, x_2) = \|\mathbf{z}_{x_1}\sqrt{\mathbf{k}} - \mathbf{z}_{x_2}\sqrt{\mathbf{k}}\|_2^2 = \|(\mathbf{z}_{x_1} - \mathbf{z}_{x_2})^2\mathbf{k}\|_1 = \sum_{u \in U} (z(u+x_1) - z(u+x_2))^2 \mathbf{k}(u) \tag{4.1}$$

with $\mathbf{k}$ being a non-negative windowing kernel defined over the neighbourhood U. Typically, $\mathbf{k}$ is rotationally symmetric and the weights $\mathbf{k}(u)$ are decided by the spatial distance from the center. In the original paper on NL-means, the authors recommend using a Gaussian function (with a fixed standard deviation) as the windowing kernel $\mathbf{k}$, as given in [43].

In this thesis, to establish non-local methods with foveation, we replace the windowed distance $d(x_1,\, x_2)$ with the foveated distance:

$$d^{\mathrm{FOV}}(x_1,\, x_2) = \|\mathbf{z}_{x_1}^{\mathrm{FOV}} - \mathbf{z}_{x_2}^{\mathrm{FOV}}\|_2^2 \tag{4.2}$$

where $\mathbf{z}_x^{\mathrm{FOV}} : U \to \mathbb{R}$ is a foveated patch obtained by foveating the image z at the fixation point x. This foveation is accomplished through a specially designed patch foveation operator $\mathcal{F}$:

$$\mathbf{z}_x^{\mathrm{FOV}}(u) = \mathcal{F}[z,\, x](u) \qquad u \in U \tag{4.3}$$

where $\mathcal{F}[\,\cdot\,,\, x]$, for $x \in X$, works as a spatially variant blurring operator with decreasing bandwidth (i.e. increasing blur) as we move outwards from the fixation point x. Strictly speaking, $\mathbf{z}_x^{\mathrm{FOV}}(u)$ gets progressively blurrier as $|u|$ increases. Similarly, we state the noise-free foveated patch to be $\mathbf{y}_x^{\mathrm{FOV}}$.

The noisy patches $\mathbf{z}_x$ follow a non-central χ^2 distribution and, thus, we can compute the mathematical expectation $E\{\,\cdot\,\}$ of the distance operator d defined in Equation (4.1):

$$\begin{aligned} E\{d(x_1,x_2)\} &= E\left\{\|(\mathbf{z}_{x_1} - \mathbf{z}_{x_2})^2\mathbf{k}\|_1\right\} = \|E\left\{(\mathbf{z}_{x_1} - \mathbf{z}_{x_2})^2\right\}\mathbf{k}\|_1 = \\ &= \|\left((\mathbf{y}_{x_1} - \mathbf{y}_{x_2})^2 + 2\sigma^2\right)\mathbf{k}\|_1 = \|(\mathbf{y}_{x_1} - \mathbf{y}_{x_2})^2\mathbf{k}\|_1 + 2\sigma^2 \sum_{u\in U} \mathbf{k}(u) \end{aligned} \tag{4.4}$$

Due to the constrained design of foveation operators, the foveated distance d^{FOV} induced by the associated $\mathcal{F}$ is, in terms of expectation under zero-mean i.i.d. white Gaussian noise, guaranteed to be equivalent to the patch distance d induced by the corresponding windowing kernel $\mathbf{k}$.

In principle, under the condition of cautious design, the modification of the distance d to d^{FOV} can be implemented to any non-local method based on pairwise patch comparision. To endorse the quantitative effectiveness of the foveated self-similarity as a regularization *prior* for natural images, we scrutinize the NL-means image denoising problem.

4.1 Foveated Non-local Means Denoising

The foveated self-similarity can be leveraged in a number of imaging applications. The removal of additive white Gaussian noise is the most ubiquitously addressed application for assessing the efficacy of any descriptive (or, generative) model of natural images. The Foveated NL-means modifies the classical NL-means denoising filter by computing the averaging weights based on the foveated patch distance instead of the conventional windowed patch distance.

4.1.1 Constraints on Foveation Operator

Our goal is to design a patch foveation operator $\mathcal{F}$ suitable for replacing the windowing distance used in the pairwise comparision of patches in non-local algorithms. We observe that in the ideal case of perfect non-local similarity, where $\mathbf{y}_{x_1}$ and $\mathbf{y}_{x_2}$ are identical, the expectation in Equation (4.4) reduces to:

$$E\{d(x_1,\, x_2)\} = 2\sigma^2 \|\mathbf{k}\|_1 \tag{4.5}$$

This simple equality plays an important role in the development of the foveation framework. The following constraints are imposed on $\mathcal{F}$:

Linearity: $\mathcal{F}$ is linear with respect to the image and translation invariant with respect to the image domain $X \subset \mathbb{Z}^2$, i.e. for an arbitrary pair of images $z_1,\, z_2$ with fixation point $x \in X$:

$$\mathcal{F}[\lambda_1 z_1 + \lambda_2 z_2,\, x - \tau] = \lambda_1 \mathcal{F}[z_1(\,\cdot\, + \tau),\, x] + \lambda_2 \mathcal{F}[z_2(\,\cdot\, + \tau),\, x] \tag{4.6}$$

for any $\lambda_1, \lambda_2 \in \mathbb{R}$, and $\tau \in \mathbb{Z}^2$. This translation invariance implies that if we translate both the image and the fixation point by a shift τ, then the foveated patch does not change.

Non-negativity: For a non-negative image, the foveated patches are always non-negative, i.e.

$$\text{if } z(x) \geqslant 0 \quad \forall x \in X, \text{ then } \mathcal{F}[z,\, x](u) \geqslant 0 \quad \forall u \in U,\ \forall x \in X. \tag{4.7}$$

Central Acuity: Foveated patches are fully sharp at their center, i.e.

$$\exists \alpha > 0 : \mathcal{F}[z,\, x](0) = \alpha\, z(x) \qquad \forall x \in X. \tag{4.8}$$

This property aims at mimicking the peak of the visual acuity at the fovea, as illustrated in Fig. 3.1. The constant α is a crucial design parameter of the foveation operator and its value will be determined in Sec. 4.1.2.

Flat-field Preservation: $\mathcal{F}$ maps a flat image onto flat patches, i.e.

$$\begin{aligned} &\exists \alpha > 0 : \forall c > 0 \quad \text{if} \quad z(x) = c\,, \quad \text{then} \\ &\mathcal{F}[z,\, x](u) = \alpha\, c \qquad \forall u \in U, \quad \forall x \in X. \end{aligned} \tag{4.9}$$

While this property appears intuitive, it is striking how seldom it is verified in the inner computations of image processing algorithms [4]. For example, the multiplication against a non-uniform windowing kernel $\mathbf{k}$, as in Eq. (4.1), prevents this property.

Compatibility: The property asserts that in the ideal case when the noise-free foveated patches are perfectly identical, the mathematical expectation of the corresponding foveated distance equals the expectation of the windowed distance, as given in Equation (4.5), i.e.

$$\text{if } \mathbf{y}_{x_1}^{\mathrm{FOV}} = \mathbf{y}_{x_2}^{\mathrm{FOV}}, \text{ then } E\left\{d^{\mathrm{FOV}}(x_1, x_2)\right\} = E\left\{\|(\mathbf{z}_{x_1}^{\mathrm{FOV}} - \mathbf{z}_{x_2}^{\mathrm{FOV}})\|_2^2\right\} = 2\sigma^2\|\mathbf{k}\|_1 \tag{4.10}$$

where $\mathbf{z}_{x_1}^{\mathrm{FOV}}$, $\mathbf{z}_{x_2}^{\mathrm{FOV}}$ are the foveated patches, as defined in Equation (4.3).

Compatibility is the most important of all the requirements because it allows using d^{FOV} as a direct replacement of d, without the need to modify any other parameters, e.g. tuning parameter h.

4.1.2 Construction of Foveation Operator

We construct the foveation operator $\mathcal{F}$, which satisfies the above requirements, by adjusting the scaling and spread of the blur PSFs in such a way that their ℓ^1 -norm is constant and their squared ℓ^2 -norm equals the corresponding value of the windowing kernel $\mathbf{k}$. The PSFs are denoted as $\{v_u\}_{u\in U}$.

Linearity and Non-negativity: An operator $\mathcal{F}$ satisfies the linearity requirement iff it can be expressed as:

$$\mathbf{z}_x^{\mathrm{FOV}}(u) = \mathcal{F}[z,\, x](u) = \sum_{\xi\in\mathbb{Z}^2} z(\xi + x)v_u(\xi - u) \qquad \forall u \in U \tag{4.11}$$

which extends from the definition of (4.3). This implies that the pixel at position u in the foveated patch $\mathbf{z}_x^{\mathrm{FOV}}$ is obtained by applying a specific kernel v_u to the neighbourhood $z(x+u)$. To satisfy non-negativity, the kernels must be $v_u \geqslant 0$. Thus, as a consequence of Equation (4.11), the foveation operator $\mathcal{F}$ is completely determined by the collection of PSFs $\{v_u\}_{u\in U}$.

Central Acuity: Central acuity holds iff v_0 is a scaled discrete Dirac impulse having value $\alpha > 0$, i.e.

$$v_0(x) = \begin{cases} \alpha & \text{if } x = 0 \\ 0 & \text{if } x \neq 0 \end{cases} \tag{4.12}$$

Flat-field Preservation: This constraint helps determine the constant ℓ^1 -norm. From Equations (4.8) and (4.11), we obtain:

$$\alpha = \sum_{\xi \in \mathbb{Z}^2} v_u(\xi - u) = \sum_{\xi \in \mathbb{Z}^2} v_u(\xi) = \|v_u\|_1 \qquad \forall u \in U \tag{4.13}$$

i.e. all the PSFs $\{v_u\}, u \in U$, have the same ℓ^1 -norm equal to α.

Compatibility: This constraint helps determine the squared ℓ^2 -norm condition. If $\mathcal{F}[y,\, x_1] = \mathbf{y}^{\text{FOV}}_{x_1} = \mathbf{y}^{\text{FOV}}_{x_2} = \mathcal{F}[y,\, x_2]$, then, because of the linearity of $\mathcal{F}$:

$$\begin{aligned} E\left\{d^{\text{FOV}}(x_1, x_2)\right\} &= E\|\mathcal{F}[z, x_1] - \mathcal{F}[z, x_2]\|_2^2 = \\ &= E\|\mathbf{y}^{\text{FOV}}_{x_1} - \mathbf{y}^{\text{FOV}}_{x_2} + \mathcal{F}[\eta, x_1] - \mathcal{F}[\eta, x_2]\|_2^2 \\ &= E\|\mathcal{F}[\tilde{\eta}, x_1]\|_2^2 \end{aligned} \tag{4.14}$$

where $\tilde{\eta}(\cdot) \sim \mathcal{N}(0, 2\sigma^2)$. Following this line of deduction, we have:

$$\begin{aligned} E\left\{d^{\text{FOV}}(x_1, x_2)\right\} &= E\left\{\sum_{u \in U} \mathcal{F}^2[\tilde{\eta}, x_1](u)\right\} = \sum_{u \in U} E\left\{\mathcal{F}^2[\tilde{\eta}, x_1](u)\right\} = \\ &= \sum_{u \in U} \text{var}\left\{\mathcal{F}[\tilde{\eta}, x_1](u)\right\} = \sum_{u \in U} \text{var}\left\{\sum_{\xi \in \mathbb{Z}^2} \tilde{\eta}\left(\xi + x_1\right) v_u\left(\xi - u\right)\right\} \\ &= \sum_{u \in U} \sum_{\xi \in \mathbb{Z}^2} \text{var}\left\{\tilde{\eta}\left(\xi + x_1\right) v_u\left(\xi - u\right)\right\} = \sum_{u \in U} \sum_{\xi \in \mathbb{Z}^2} 2\sigma^2 v_u^2(\xi) = \\ &= 2\sigma^2 \sum_{u \in U} \|v_u\|_2^2 = 2\sigma^2 \sum_{u \in U} \mathbf{k}(u) \end{aligned}$$

Therefore, the compatibility requirement is satisfied when the sum of the squared ℓ^2 -norms of all the PSFs of $\mathcal{F}$ equals the ℓ^1 -norms of the kernel:

$$\sum_{u \in U} \|v_u\|_2^2 = \|\mathbf{k}\|_1 = \sum_{u \in U} \mathbf{k}(u) \tag{4.15}$$

Pixel-wise ℓ^2 -norm condition: The series of aforementioned equalities reveal that a stricter *pixel-wise* compatibility holds provided that the squared ℓ^2 -norms of all the PSFs (i.e. for each $u \in U$) coincide with the corresponding value of the windowing kernel:

$$\|v_u\|_2^2 = \mathbf{k}(u) \qquad \forall u \in U \tag{4.16}$$

By considering the above equality in the case $u = 0$, then from Equation (4.12) we obtain that $\|v_0\|_2^2 = \alpha^2 = \mathbf{k}(0)$, which determines the value of the constant α appearing in the central acuity and flat-field preservation constraints as:

$$\alpha = \sqrt{\mathbf{k}(0)}$$

To summarize, the foveation operator $\mathcal{F}$ satisfies the five constraints with pixel-wise compatibility iff:

$$\mathcal{F}[z,\, x](u) = \sum_{\xi \in \mathbb{Z}^2} z(\xi + x) v_u(\xi - u) \qquad \forall u \in U \tag{4.17}$$

with non-negative PSFs $\{v_u\}, u \in U$, such that:

$$\|v_u\|_1 = \sqrt{\mathbf{k}(0)} > 0 \qquad \forall u \in U \tag{4.18}$$

$$\|v_u\|_2^2 = \mathbf{k}(u) \qquad \forall u \in U \tag{4.19}$$

$$v_0 \text{ is a discrete Dirac impulse having value } \sqrt{\mathbf{k}(0)} \tag{4.20}$$

4.1.3 Gaussian Foveation Operators

Taking the central-limit principle into consideration, we argue that Gaussian distributions well approximate the final blurring effect of space-variant processes involved in foveation, as described in Section 4.1, thus providing a legitimate model for the PSFs $\{v_u\}_{u \in U}$. We present the construction of foveation operators based on a family of Gaussian PSFs, such that all the PSFs share the same ℓ^1 -norm, given by Equation (4.18), and have ℓ^2 -norms given by Equation (4.19).

Isotropic Foveation Operator: This is the case for circular symmetric PSFs $\{v_u\}_{u \in U}$, as they attenuate image features regardless of the feature's orientation. The attenuation strength depends only on the distance $|u|$ from the patch center, given a windowing kernel $\mathbf{k}$ (as in the *isotropic* case). We define g_ς as the circularly-symmetric bivariate Gaussian probability density function (PDF) with zero-mean and diagonal covariance matrix $\mathbf{\Sigma}_\varsigma$:

$$g_\varsigma(\xi) = \frac{1}{2\pi\varsigma^2} \exp\left(-\frac{|\xi|^2}{2\varsigma^2}\right) \qquad \xi \in \mathbb{R}^2 \tag{4.21}$$

The standard deviation parameter ς determines the spread of the Gaussian PDF. Using a bit of calculus shows:

$$\|g_\varsigma\|_1 = 1\,, \qquad \|g_\varsigma\|_2^2 = \frac{1}{4\pi\varsigma^2} \tag{4.22}$$

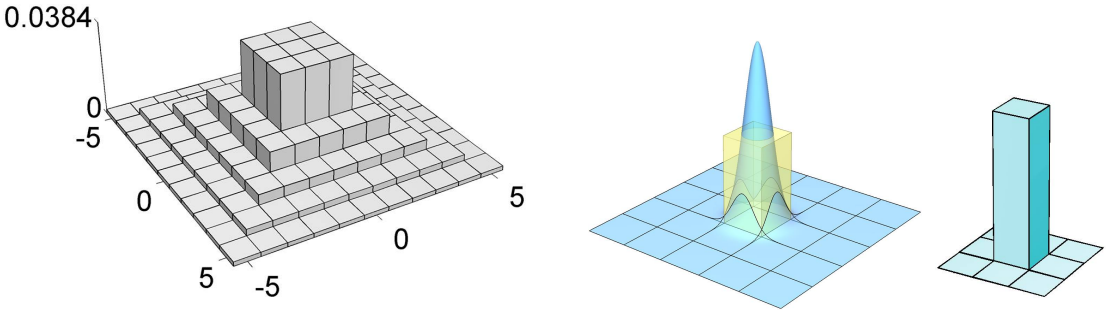
Hence, to obtain a PSF v_u that satisfies Equations (4.18) and (4.19), we first scale g_ς by multiplying it with $\alpha = \sqrt{\mathbf{k}(0)}$ and then dilating it by choosing a standard deviation parameter ς that solves $\|\sqrt{\mathbf{k}(0)}\, g_\varsigma\|_2^2 = \mathbf{k}(u)$, i.e.

$$\varsigma = \frac{1}{2\sqrt{\pi}}\sqrt{\frac{\mathbf{k}(0)}{\mathbf{k}(u)}} \tag{4.23}$$

As a consequence of the pixel-wise compatibility, the above Equation installs a direct link between the spread of the PSF v_u and the value of the windowing kernel $\mathbf{k}(u)$. When $\mathbf{k}(u)$ is small (i.e. at the periphery of the patch), the blur caused by the PSF is large, thus mimicking the effects of foveation. Therefore, using the value of ς calculated in Equation (4.23), we define:

$$v_u(\xi) = \sqrt{\mathbf{k}(0)}\, g_\varsigma(\xi) = \frac{2\mathbf{k}(u)}{\sqrt{\mathbf{k}(0)}} \exp\left(-2\pi|\xi|^2 \left(\frac{\mathbf{k}(u)}{\mathbf{k}(0)}\right)\right) \qquad \xi \in \mathbb{R}^2 \tag{4.24}$$

The central-acuity constraint is achieved by "manually" re-defining v_0, according to Eq. (4.12), with $\alpha = \sqrt{\mathbf{k}(0)}$. This value is an exception from the general form in Eq. (4.24) for $u = 0$ (which is, $v_0(\xi) = 2\sqrt{\mathbf{k}(0)}\exp(-2\pi|\xi|^2)$). However, the numerical difference between the discrete Dirac impulse value and the discrete Gaussian representation v_0 is negligible, as in Fig 4.1c.



(a) Windowing kernel [32]. (b) (c)

Figure 4.1: (a) A windowing kernel $\mathbf{k}$ of size 11×11 used in the computation of the similarity weights in the NL-means. (b) Scaled discrete Dirac impulse. (c) Gaussian PSF after discretization (Reproduced from [35]).

Anisotropic Foveation Operator: These operators generalize the isotropic PSFs $\{v_u\}$ by utilizing an elliptical Gaussian PDF $g_\varsigma^{\rho,\,\vartheta}$ whose covariance matrix depends not only on $\varsigma > 0$, but also on a parameter $\rho > 0$ that determines the elongation of the PDF, and on an angular parameter $\vartheta \in \mathbb{R}$ that controls the orientation of the elliptical PDF. Specifically,

$$\mathbf{\Sigma}_\varsigma^{\rho,\,\vartheta} = \varsigma_\vartheta^2\, \mathbf{R}_\vartheta\, \mathbf{D}_\rho\, \mathbf{R}_\vartheta^T \tag{4.25}$$

where the diagonal matrix $\mathbf{D}_\rho = \begin{bmatrix} \rho & 0 \\ 0 & 1/\rho \end{bmatrix}$ determines the PDF elongation and $\mathbf{R}_\vartheta = \begin{bmatrix} \cos(\vartheta) & -\sin(\vartheta) \\ \sin(\vartheta) & \cos(\vartheta) \end{bmatrix}$ is a rotation matrix of angle ϑ :

$$g_\varsigma^{\rho,\,\vartheta}(\xi) = \frac{1}{2\pi\varsigma^2} \exp\left(-\frac{1}{2}\, \boldsymbol{\xi}^T\, \left(\mathbf{\Sigma}_\varsigma^{\rho,\,\vartheta}\right)\, \boldsymbol{\xi}\right) \qquad \xi \in \mathbb{R}^2 \tag{4.26}$$

where $\boldsymbol{\xi}$ is a column-vector representation of ξ . Clearly, $\rho = 1$ corresponds to the circularly-symmetric case, and it follows that $g_\varsigma^{1,\,\vartheta} = g_\varsigma$ for any $\vartheta \in \mathbb{R}$ and $\varsigma > 0$.

We observe that the PDF $g_\varsigma^{\rho,\,\vartheta}$ conforms to the same norm as the circularly symmetric g_ς , i.e.

$$\|g_\varsigma^{\rho,\,\vartheta}\|_1 = 1\,, \qquad \|g_\varsigma^{\rho,\,\vartheta}\|_2^2 = \frac{1}{4\pi\varsigma^2} \qquad \forall\rho > 0,\ \ \forall\vartheta \in \mathbb{R} \tag{4.27}$$

This fact is guaranteed by the definition of $\mathbf{D}_\rho$, which has an unitary determinant. Also, the constraints in Section 4.1.1 are all satisfied, and their proof depends on the fact that the norms are invariant with respect to elliptical deformation of the PDF $g_\varsigma^{\rho,\,\vartheta}$.

Hence, to construct the *anisotropic* foveation operator, as in the isotropic case, we first scale $g_\varsigma^{\rho,\,\vartheta}$ by multiplying it with $\sqrt{\mathbf{k}(0)}$ and then dilating it by choosing the standard deviation as:

$$\varsigma = \frac{1}{2\sqrt{\pi}} \sqrt{\frac{\mathbf{k}(0)}{\mathbf{k}(u)}} \tag{4.28}$$

same as in Equation (4.23). We observe that the conditions stated in Equations (4.18) and (4.19) are met for any combination of ρ and ϑ, without compromising the validity of the ℓ^1 and ℓ^2 -norms. Models of acuity in the HVS suggests that ρ depends on $|u|$, and ϑ on $\angle u$, where $u \in U$.

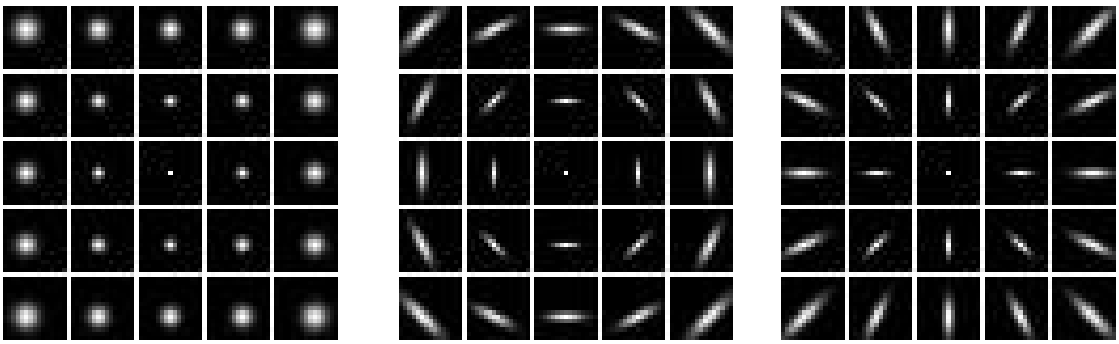
We focus our attention on a specific simplified design, where ρ is constant, and where $\vartheta = \angle u + \theta$, $\theta \in \mathbb{R}$ being an angular offset. This choice leads to a class of anisotropic foveation operators:

$$\mathcal{F}_{\rho,\,\theta}[z, x](u) = \sum_{\xi \in \mathbb{Z}^2} z(\xi + x) v_u^{\rho,\,\theta}(\xi - u) \qquad \forall u \in U \tag{4.29}$$

defined through the PSFs:

$$v_u^{\rho,\,\theta} = \begin{cases} \sqrt{\mathbf{k}(0)}\; g_{\varsigma}^{\rho,\,\angle u + \theta} & u \neq 0 \\ \sqrt{\mathbf{k}(0)}\; g_{\frac{1}{2\sqrt{\pi}}} & u = 0 \end{cases} \tag{4.30}$$

The anisotropic foveation operators $\mathcal{F}_{\rho,\,\theta}$ satisfy all the five constraints in Sec. 4.1.1, for any combination of $\rho > 0$ and $\theta \in \mathbb{R}$. When $\theta = 0$ and $\rho > 1$, the PSFs yield a *radial* foveation operator. Conversely, if $\theta = \pi/2$ and $\rho > 1$, the PSFs yield a *tangential* foveation operator. When $\rho = 1$, $\mathcal{F}_{\rho,\,\theta}$ coincides with the *isotropic* foveation operator (see Fig. 4.2, reproduced from [34]).



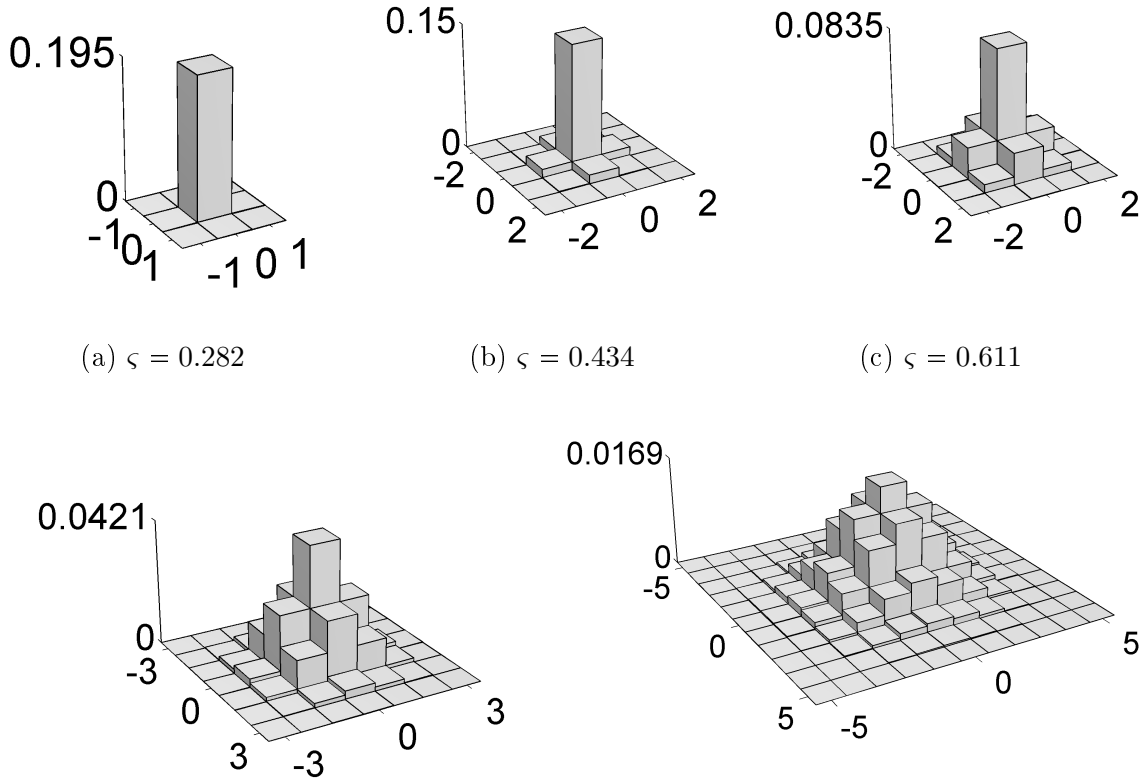
(a) Circular (b) Tangential (c) Radial

Figure 4.2: Illustration of (a) isotropic; and (b), (c) anisotropic foveation operators.

4.1.4 Illustrations of Foveation Operator

We give an illustration of the construction of a foveation operator and of its PSFs for a given windowing kernel $\mathbf{k}$. For this example, we consider the windowing kernel $\mathbf{k}$ used in the NL-means implementation by [43] for a neighbourhood U of size 11×11 pixels, as shown in Fig. 4.1a. Due to its symmetric nature, $\mathbf{k}$ takes a very limited number of distinct values $\mathbf{k}(u)$, $u \in U$, reported as 0.0384, 0.0162, 0.0082, 0.0041, 0.0017 . For each distinct value of $\mathbf{k}(u)$, Eq. (4.23) gives a distinct value of the standard deviation parameter ς and Eq. (4.24) defines the corresponding distinct PSF v_u .

The blurring kernels are shown in Fig. 4.3 and each of these kernels is based on a discrete Gaussian kernel g_ς of size $(2\lceil 3\varsigma \rceil + 1) \times (2\lceil 3\varsigma \rceil + 1)$, where $\lceil \cdot \rceil$ is the ceiling function and the radius of the kernel is $\lceil 3\varsigma \rceil$. The frequency responses (read: [32]) of the five blurring kernels all attain their maximum at the origin, and the value of this maximum is equal to $\sqrt{\mathbf{k}(0)} = 0.196$, as is implied by the Equation (4.18). The squared ℓ^2 -norm values $\|v_u\|_2^2$ for these five PSFs are 0.0379, 0.0231, 0.009, 0.0041, 0.0017, respectively, which are nearly equal to the corresponding values of $\mathbf{k}(u)$, $\forall u \in U$, as given by Equation (4.19). It should be noted that the minor difference is due to the discretization of the Gaussian PDF g_ς used in our implementation, while the ℓ^2 -norm condition in Equation (4.27) assumes continuous domain variables.



(a) $\varsigma = 0.282$ (b) $\varsigma = 0.434$ (c) $\varsigma = 0.611$

(d) $\varsigma = 0.861$ (e) $\varsigma = 1.360$

Figure 4.3: The five blurring kernels, corresponding to the five unique values of the window $\mathbf{k}$ (Reproduced from [35]).

In [36], a clear analogy was shown between the orientation preference in the HVS, and the angular position and orientation of the radial receptors, i.e. $\theta = 0$. In this sense, radial foveation induces a patch similarity measure that mimics the HVS sensitivity. This relation substantiates why anisotropic patch foveation improves the denoising performance with respect to the isotropic foveation.

4.1.5 Experimental Results, and Discussion

The PSNR and SSIM values for various experiments reported in [36], [61] show that Foveated NL-means is highly successful at image denoising as compared to standard NL-means, particularly at high noise levels. The results indicate that the foveated distance is a more effective measure for assessing the non-local self-similarity in images. The superior sharpness and contrast achieved by foveation is a result of high frequency content at the periphery (of the patch) enjoying a much shorter-range correlation than the low frequency content at the center.

Noisy Image Denoised Image

Figure 4.4: Illustration of FNLM.

5. FOVEATED NL-MEANS FOR COLOR IMAGES

In this chapter, we extend the Foveated NL-means (FNLM) denoising algorithm to color images, and introduce a cross-channel paradigm to exploit the correlation between color information. The cross-channel kernel follows from the spatially-variant nature of color perception in the HVS, which suggests that peripheral color vision is inferior in comparision to foveal color vision (described in Sec. 3.4).

The FNLM method can be developed for color images by building upon the foveation operator constraints described for grayscale images, incorporating the color channels in further definitions, and exploiting the correlation between channels by introducing a unified color-mixing foveation operator. Denoising is performed in the RGB color space without resorting to a color-space transformations.

5.1 Preliminaries

For grayscale images, the observation model is:

$$z(x) = y(x) + \eta(x) \qquad x \in X \subset \mathbb{Z}^2 \tag{5.1}$$

with the variables having the same interpretation as mentioned in Section 1.2. When considering color images, it is common practice to assume full-color (or, demosaiced) images. We model a noisy RGB image as:

$$z_{\mathrm{RGB}}(x) = y_{\mathrm{RGB}}(x) + \eta_{\mathrm{RGB}}(x) \qquad x \in X \subset \mathbb{Z}^2 \tag{5.2}$$

where $y_{\mathrm{RGB}} = [y_{\mathrm{R}}, y_{\mathrm{G}}, y_{\mathrm{B}}]$ is the true underlying image, and $\eta_{\mathrm{RGB}} = [\eta_{\mathrm{R}}, \eta_{\mathrm{G}}, \eta_{\mathrm{B}}]$ is the independent white Gaussian noise, where $\eta_{\mathrm{c}}(\,\cdot\,) \sim \mathcal{N}(0, \sigma_{\mathrm{c}}^2)$ for $\mathrm{C} \in \{\mathrm{R},\ \mathrm{G},\ \mathrm{B}\}$; the variances are assumed to be the same, i.e. $\sigma_{\mathrm{c}}^2 = \sigma^2$.

The foveated distance for color images is defined as:

$$d_{\mathrm{j}}^{\mathrm{FOV}}(x_1, x_2) = \|\mathbf{z}_{x_1,\mathrm{j}}^{\mathrm{FOV}} - \mathbf{z}_{x_2,\mathrm{j}}^{\mathrm{FOV}}\|_2^2 \qquad \mathrm{j} \in \mathrm{C} \in \{\mathrm{R},\ \mathrm{G},\ \mathrm{B}\} \tag{5.3}$$

where $\mathbf{z}_{x,\mathrm{j}}^{\mathrm{FOV}} : \{U \times \{\mathrm{R},\ \mathrm{G},\ \mathrm{B}\}\} \to \mathbb{R}$ is a foveated patch obtained by foveating the color image z_{RGB} at the fixation point x in $u \in U$. Hence, we define the color foveation operator $\mathcal{F}_{\mathrm{RGB}}$, which integrates the blurring PSFs and color-mixing, as:

$$\begin{aligned}\mathbf{z}^{\mathrm{FOV}}_{x,\,\mathrm{j}}(u) = \mathcal{F}_{\mathrm{RGB}}[z_{\mathrm{RGB}}, x, \mathrm{j}](u) &= \sum_{\mathrm{i}\in C} \mathcal{M}_{\mathrm{i},\,\mathrm{j}}(u)\mathcal{F}[z^i,\, x](u) \qquad \mathrm{i},\mathrm{j} \in \mathrm{C} \in \{\mathrm{R},\, \mathrm{G},\, \mathrm{B}\} \\ &= \sum_{\mathrm{i}\in C} \mathcal{M}_{\mathrm{i},\,\mathrm{j}}(u) \sum_{\xi\in\mathbb{Z}^2} z^i(\xi + x) v_u(\xi - u)\end{aligned} \tag{5.4}$$

where $\mathcal{M}_{\mathrm{i},\,\mathrm{j}}(u)$ is a real number for a fixed $\mathrm{i},\mathrm{j} \in \mathrm{C}$, $u \in U$, and z^i are the individual noisy channels of a color image z_{RGB}. Also, $\mathcal{F}$ is identical over all the channels as the windowing kernel $\mathbf{k}$, for $u \in U$, is same over all the color channels.

5.2 Cross-channel Paradigm

In [38], it was shown that color channels are correlated. The structure of patches, $\mathbf{z}^{\mathrm{FOV}}_{x,\,\mathrm{j}}$, in the same location, remains roughly the same across all color channels. With this in mind, our FNLM method for color images utilizes these cross-color dependencies, instead of a naive application of FNLM to the color channels separately. As a consequence, we get an improvement in the denoising performance.

We introduce a color-mixing array $\mathcal{M}_{\mathrm{i},\,\mathrm{j}}$, which consists of color-mixing kernels, given by $\mathcal{D}$, and its complement $\tilde{\mathcal{D}}$, defined over $u \in U$, as:

$$\mathcal{M}_{\mathrm{i},\,\mathrm{j}} = \begin{cases} \mathcal{D} & \text{if } i = j \\ \tilde{\mathcal{D}} & \text{if } i \neq j \end{cases}, \text{where } \mathrm{i},\mathrm{j} \in \mathrm{C} \in \{\mathrm{R},\, \mathrm{G},\, \mathrm{B}\}$$

The array operates in the 3- channel color space and is defined over the neighbourhood $u \in U$. We impose the following constraints on the kernel, $\mathcal{D}$:

i. The value of each pixel of the kernel lies in the range $[0.33,\, 1]$.

ii. The size of the kernel is same as the size of the discrete Gaussian kernel.

iii. The ℓ^1 -norm of the kernel is calculated channel-wise, and not pixel-wise.

iv. For an n -channel image, the net sum over all channels of $\mathcal{D}$ and $\tilde{\mathcal{D}}$ is 1 , i.e.

$$\mathcal{D} + (n-1)\,\tilde{\mathcal{D}} = 1 \tag{5.5}$$

The above equation states that the color-mixing kernel is applied successively to the individual channels, with the complement being applied to the remaining $(n-1)$ channels. Given the condition in Eq. (5.5), we define $\tilde{\mathcal{D}}$ as:

$$\tilde{\mathcal{D}} = \frac{1 - \mathcal{D}}{n-1} \tag{5.6}$$

where $n = 3$ for an RGB image.

The $n \times n$ color-mixing matrix, defined by $\mathcal{D}$ and $\tilde{\mathcal{D}}$ kernels, for $u \in U$, is:

$$\mathcal{M}(u) = \begin{bmatrix} \mathcal{D}(u) & \tilde{\mathcal{D}}(u) & \cdots & \tilde{\mathcal{D}}(u) \\ \tilde{\mathcal{D}}(u) & \mathcal{D}(u) & \cdots & \tilde{\mathcal{D}}(u) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathcal{D}}(u) & \tilde{\mathcal{D}}(u) & \cdots & \mathcal{D}(u) \end{bmatrix}$$

We can now formally define the three-dimensional unified PSFs:

$$V_u = \mathcal{M}_{\mathrm{i,\,j}}(u)\, v_u(\xi) \qquad \forall u \in U,\ \forall \xi \in \mathbb{Z}^2 \tag{5.7}$$

where i, j $\in$ C $\in$ {R, G, B} and $\mathcal{M}(u) : \{\mathrm{C} \times \mathrm{C}\} \to \mathbb{R}$. The unified operator utilizes the idea that underlying image structures (e.g. objects, edges, details, etc) are the same across all color channels. The family of PSFs have a constant ℓ^1 -norm, and an ℓ^2 -norm dependent on the value of $\mathbf{k}(u)$.

5.3 Constrained Design of the Unified Operator

We investigate the effects of the unified color-mixing foveation operator on the statistical characteristics of the noisy color image, especially the expectation and variance. All the constraints on the foveation operator mentioned in Section 4.1.1 are applicable for color images, with one addendum - Linear Separability. The constraints and mathematical expectation of the corresponding foveated operator, is extended to account for the color-mixing operator, for i, j $\in$ C $\in$ {R, G, B}.

Linear Separability: The ℓ^p -norm of the unified PSFs $\{V_u, u \in U\}$ is the product of the ℓ^p -norm of the foveation operator and the ℓ^p -norm of the color-mixing operator, taken channel-wise, i.e.

$$\|V_u\|_{p,\,\mathrm{i}} = \|\mathcal{M}(u)\, v_u\|_{p,\,\mathrm{i}} = \|\mathcal{M}(u)\|_{p,\,\mathrm{i}}\|\, v_u\|_p \tag{5.8}$$

where **i** is the color channel under consideration. Similarly,

$$\begin{aligned} \|V_u\|^2_{2,\mathrm{i}} = \|\mathcal{M}(u)\, v_u\|^2_{2,i} &= \|\mathcal{M}(u)\|^2_{2,\mathrm{i}}\, \|v_u\|^2_2 = \\ &= \mathbf{k}(u) \left(D^2(u) + \tilde{D}^2(u) + \overset{(n-3)}{\cdots} + \tilde{D}^2(u) \right) \\ &= \mathbf{k}(u) \left(\mathcal{D}^2(u) + \frac{(1-\mathcal{D}(u))^2}{n-1} \right) \end{aligned} \tag{5.9}$$

Hence, the constraint is satisfied when the sum of the squared ℓ^2 -norms of all the PSFs $\{V_u\}$ of $\mathcal{F}_{\mathrm{RGB}}$, over the individual color channels, is given by:

$$\sum_{\mathrm{i}\in C}\sum_{u\in U}\|\mathcal{M}(u)\,v_u\|_{2,\mathrm{i}}^2=\sum_{\mathrm{i}\in C}\sum_{u\in U}\mathbf{k}(u)\left(\mathcal{D}^2(u)+\frac{(1-\mathcal{D}(u))^2}{n-1}\right) \tag{5.10}$$

Non-negativity: For a non-negative color image, the color-mixed foveated patches are always non-negative, i.e.

$$\text{if } z_{\mathrm{RGB}}(x)\geqslant 0\,,\ \text{then } \mathcal{F}_{\mathrm{RGB}}[z_{\mathrm{RGB}},x,\mathrm{j}](u)\geqslant 0 \qquad \forall u\in U,\ \forall x\in X. \tag{5.11}$$

Central Acuity: Color-mixed foveated patches are fully sharp at their center and holds iff v_0 is a scaled discrete Dirac impulse, i.e.

$$\exists\beta>0:\mathcal{F}_{\mathrm{RGB}}[z_{\mathrm{RGB}},x,\mathrm{j}](0)=\beta\,z_{\mathrm{RGB}}(x) \qquad \forall x\in X. \tag{5.12}$$

Flat-field Preservation: $\mathcal{F}_{\mathrm{RGB}}$ maps a flat color image onto flat patches, i.e. $z_{\mathrm{RGB}}(x)=f$. This constraint helps determine the ℓ^1 -norm. From Eq. (5.8):

$$\exists\beta>0:\mathcal{F}_{\mathrm{RGB}}[z_{\mathrm{RGB}},x,\mathrm{j}](u)=\beta\,f \qquad \forall u\in U,\ \forall f>0,\ \forall x\in X. \tag{5.13}$$

we obtain:

$$\begin{aligned}\beta=\sum_{\mathrm{i}\in C}\sum_{\xi\in\mathbb{Z}^2}\mathcal{M}_{\mathrm{i,j}}(u)v_u(\xi-u)&=\sum_{\mathrm{i}\in C}\sum_{\xi\in\mathbb{Z}^2}\mathcal{M}_{\mathrm{i,j}}(u)v_u(\xi)\\&=\|\mathcal{M}(u)\,v_u\|_{1,\mathrm{i}}\quad\forall u\in U\end{aligned} \tag{5.14}$$

i.e. all the PSFs $\{V_u\}, u\in U$, have the same ℓ^1 -norm equal to β. This norm is constant and independent of the value of $\mathbf{k}(u)$.

Compatibility: In the case of perfect self-similarity, where $\tilde{\eta}(\cdot)\sim\mathcal{N}(0,2\sigma^2)$ as in Equation (4.14), we have, due to linearity:

$$\begin{aligned}E\left\{d_{\mathrm{j}}^{\mathrm{FOV}}(x_1,x_2)\right\}&=E\|\mathcal{F}_{\mathrm{RGB}}[z_{\mathrm{RGB}},x_1,\mathrm{j}]-\mathcal{F}_{\mathrm{RGB}}[z_{\mathrm{RGB}},x_2,\mathrm{j}]\|_2^2=\\&=E\|\mathbf{y}_{x_1,\mathrm{j}}^{\mathrm{FOV}}-\mathbf{y}_{x_2,\mathrm{j}}^{\mathrm{FOV}}+\mathcal{F}_{\mathrm{RGB}}[\eta,x_1,\mathrm{j}]-\mathcal{F}_{\mathrm{RGB}}[\eta,x_2,\mathrm{j}]\|_2^2\\&=E\|\mathcal{F}_{\mathrm{RGB}}[\tilde{\eta},x_1,\mathrm{j}]\|_2^2\end{aligned} \tag{5.15}$$

We further extend this definition of expectation of the foveated distance for color images, by utilizing the ℓ^2 -norm condition in Eq. (5.10), as:

$$
\begin{aligned}
E\left\{d_{\mathrm{j}}^{\mathrm{FOV}}(x_1,x_2)\right\} &= \sum_{u\in U} E\left\{\mathcal{F}^2_{\mathrm{RGB}}[\tilde{\eta},x_1,\mathrm{j}](u)\right\} = \\
&= \sum_{u\in U} \mathrm{var}\left\{\mathcal{F}_{\mathrm{RGB}}[\tilde{\eta},x_1,\mathrm{j}](u)\right\} = \\
&= \sum_{u\in U} \mathrm{var}\left\{\sum_{\mathrm{i}\in C}\mathcal{M}_{\mathrm{i,j}}(u)\sum_{\xi\in\mathbb{Z}^2}\tilde{\eta}^{\,i}\left(\xi+x_1\right)v_u\left(\xi-u\right)\right\} = \\
&= \sum_{u\in U}\sum_{\mathrm{i}\in C}\sum_{\xi\in\mathbb{Z}^2} \mathrm{var}\left\{\tilde{\eta}^{\,i}\left(\xi+x_1\right)v_u\left(\xi-u\right)\mathcal{M}_{\mathrm{i,j}}(u)\right\} = \\
&= \sum_{u\in U}\sum_{\mathrm{i}\in C}\sum_{\xi\in\mathbb{Z}^2} 2\sigma^2\mathcal{M}^2_{\mathrm{i,j}}(u)v_u^2(\xi) = \\
&= 2\sigma^2\sum_{\mathrm{i}\in C}\sum_{u\in U}\|\mathcal{M}(u)v_u\|^2_{2,\mathrm{i}} = \\
&= 2\sigma^2\sum_{\mathrm{i}\in C}\sum_{u\in U}\mathbf{k}(u)\left(\mathcal{D}^2(u)+\frac{(1-\mathcal{D}(u))^2}{n-1}\right)
\end{aligned}
\tag{5.16}
$$

where n is the cardinality of the color set C, and i, j ∈ C ∈ {R, G, B}.

Pixel-wise ℓ^2 -norm condition: The above series of equalities suggest that a stricter pixel-wise compatibility can be imposed provided that the squared ℓ^2 -norms of each PSF (i.e. for each $u \in U$) coincide with the corresponding scaled value of the normalized windowing kernel:

$$
\|\mathcal{M}(u)\,v_u\|^2_{2,\mathrm{i}} = \mathbf{k}(u)\left(\mathcal{D}^2(u)+\frac{(1-\mathcal{D}(u))^2}{n-1}\right) \tag{5.17}
$$

By considering the above equality for the case $u = 0$, along with Eq. (5.12), we determine the value of the constant β appearing in the central acuity and flat-field preservation constraints:

$$
\|V_0\|^2_{2,\mathrm{i}} = \beta^2 = \mathbf{k}(0)\left(\mathcal{D}^2(0)+\frac{(1-\mathcal{D}(0))^2}{n-1}\right) \tag{5.18}
$$

Weight Function: In adapting FNLM for color patches, the formulation of the weight function $w(x_1, x_2)$ is defined to account for the color channels. The number of color channels are used as a normalizing factor in the weights, so as to obtain results whose magnitude is comparable to those of grayscale images when using the same windowing kernel $\mathbf{k}$. The weighted averaging of the patches is done for individual channels before the final aggregation process is carried out across all the channels. If the strength of the noise affecting the different color channels is the same, i.e. $\sigma^2_{\mathrm{R}} = \sigma^2_{\mathrm{G}} = \sigma^2_{\mathrm{B}} = \sigma^2$ the above method for calculating the weights work well.

This unified approach, denominated as C-FNLM (Color-FNLM), is facilitated by the fact that the isotropic foveation operator and color-mixing operator are obtained as linearly separable transforms. This is a consequence of the fact that the foveation kernels are circular symmetric and have a bivariate Gaussian PDF.

5.3.1 Construction and Illustration

As explained in Sec. 3.5, foveation is the result of several cascaded space-variant processes. The blurring effect of these processes can be well approximated, considering the central-limit principle, using Gaussian distributions which provide a legitimate model for the foveation PSFs $\{v_u\}_{u\in U}$. Such foveation is *isotropic* because the PSFs attenuate image features regardless of the features orientation, and the attenuation strength depends only on the distance $\|u\|_2$ from the patch center.

The construction of the unified color-mixing foveation operator V_u that satisfies the aforementioned constraints is achieved by adjusting the spread of the blur PSFs such that their ℓ^1 -norm is constant and their squared ℓ^2 -norms equal the corresponding scaled values of the normalized windowing kernel $\mathbf{k}$. For the particular case of Gaussian PSFs, the unified operator is defined as:

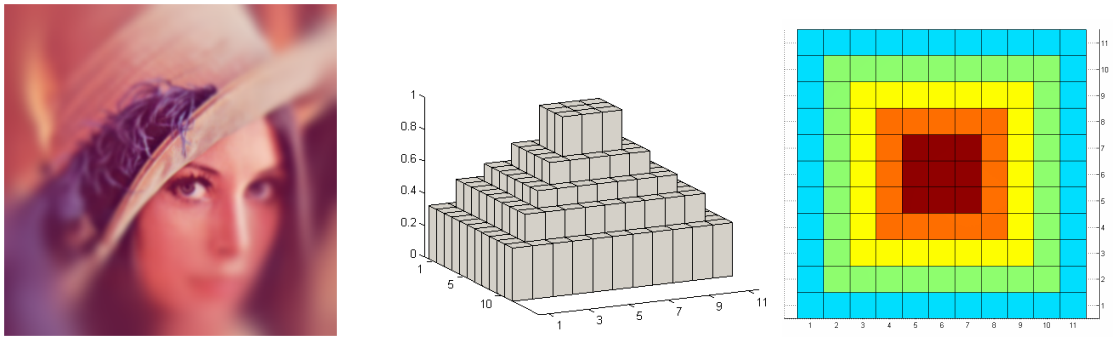
$$\begin{aligned}\mathbf{z}^{\mathrm{FOV}}_{x,\,\mathrm{j}}(u) = \mathcal{F}_{\mathrm{RGB}}[z_{\mathrm{RGB}}, x, \mathrm{j}](u) &= \sum_{\mathrm{i}\in C} \mathcal{M}_{\mathrm{i},\,\mathrm{j}}(u)\mathcal{F}[z^i,\ x](u) \qquad \mathrm{i},\mathrm{j}\in \mathrm{C}\in\{\mathrm{R},\ \mathrm{G},\ \mathrm{B}\}\\ &= \sum_{\mathrm{i}\in C} \mathcal{M}_{\mathrm{i},\,\mathrm{j}}(u) \sum_{\xi\in\mathbb{Z}^2} z^i(\xi) v_u(\xi - x - u) \qquad \forall u\in U\end{aligned} \tag{5.19}$$

where $v_u(\xi) = \sqrt{\mathbf{k}(0)}\, g_\varsigma(\xi)$ describes the bivariate Gaussian PDF with zero mean and diagonal covariance matrix, as shown in Section 4.1.3. Distinct values of the standard-deviation parameter ς defines a distinct PSF v_u , having a size of $(2\lceil 3\varsigma\rceil + 1)\times(2\lceil 3\varsigma\rceil + 1)$. The kernel's radius is $\lceil 3\varsigma\rceil$ based on a *three-sigma* rule which approximates the maximum ς of the PSF. Also, we use symmetric padding to preliminary pad z_{RGB} outside of its native domain X.

In this thesis, the color-mixing (CM) array $\mathcal{M}_{\mathrm{i},\,\mathrm{j}}$ is constructed from a variety of modified Gaussian kernels having varying standard deviation, whereas the isotropic foveation operator $\mathcal{F}$ is always constructed from the *standard* windowing kernel $\mathbf{k}$ used in the standard NL-means implementation. Due to the symmetric nature of $\mathbf{k}$, it takes a limited number of distinct values $\mathbf{k}(u)$, $u\in U$, as shown in Fig. 4.1a. After normalizing each distinct value, we rescale them as:

$$\mathcal{M}_{\mathrm{i},\,\mathrm{j}}(u) = \frac{(n-1)\,\mathbf{k}(u)}{n} + \frac{1}{n} \tag{5.20}$$

which yields distinct values of the color-mixing array, and ensures the values lies in the range $[0.33, 1]$ as shown in Fig. 5.1b. Also, $n = 3$ for RGB images.



(a) Foveated Color Image. (b) Standard CM Array. (c) Top-view of (b).

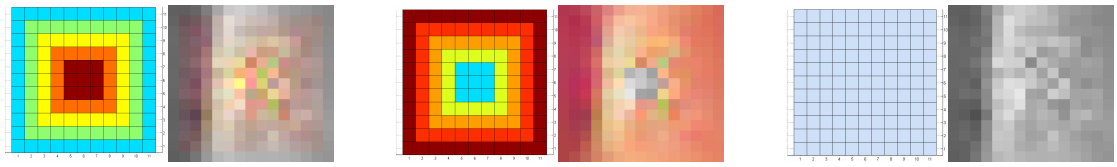
Figure 5.1: (a) Output obtained by a trivial application of FNLM on the 3−color channels. (b) Standard CM array of size 11×11. (c) Five unique values of the CM array.

For the purpose of visualization, we separately show the foveated color image $\mathcal{F}[z^i,\, x]$, the color-mixing array $\mathcal{M}_{\mathrm{i,\,j}}$, and the output of unified operator $\mathcal{F}_{\mathrm{RGB}}$. The color-mixing (CM) array constructed from a *standard* kernel, as can be seen from Fig. 5.1b, has the pixel value at the center as $\mathcal{D}(u) = 1$ and the value at the edge as $\mathcal{D}(u) = 0.33$. From this, the cross-channel paradigm dictates that the value of $\tilde{\mathcal{D}}(u)$ at the center and edge pixels are 0 and 0.33, respectively.

We also construct color-mixing arrays from kernels which do not exhibit decreasing acuity towards the periphery, i.e. *uniform* kernels and *inside-out* kernels. In the case of inside-out color-mixing array, the positions are completely reversed compared to the natural design of a standard color-mixing array, i.e. the array attains its minimum $\mathcal{D}(u) = 0.33$ at the center and increases towards the periphery to $\mathcal{D}(u) = 1$ (Fig. 5.2b). For the uniform color-mixing array, the value of $\mathcal{D}$ is uniformly 0.33 and we get the average over all the color channels (Fig. 5.2c).

5.4 Experimental Results

To assess the effectiveness of the unified operator as a regularization prior for natural color images, we quantitatively consider the denoising of noisy images. We compare the proposed algorithm with (i) NL-means algorithm for RGB images (as described in Sec. 2.1.3), and (ii) a trivial application of FNLM independently on the three color channels, without the color-mixing array (Fig. 5.1a). The denoising experiments are carried out on a set of four color test images of size 512×512, in the intensity range $[0, 255]$, namely *Barbara, Boats, Hill, Lena*. The noise-free images are shown in Fig. 5.3. These have been synthetically corrupted by 3 independent realizations of additive white Gaussian noise at different values of standard deviation $\sigma \in \{10, 20, 30, 50, 70\}$, according to the observation model in Eq. (5.2). We measure the denoising performance according to established quality assessment indicators: PSNR (dB) and the perceptual quality index SSIM [69].



(a) (b) (c)

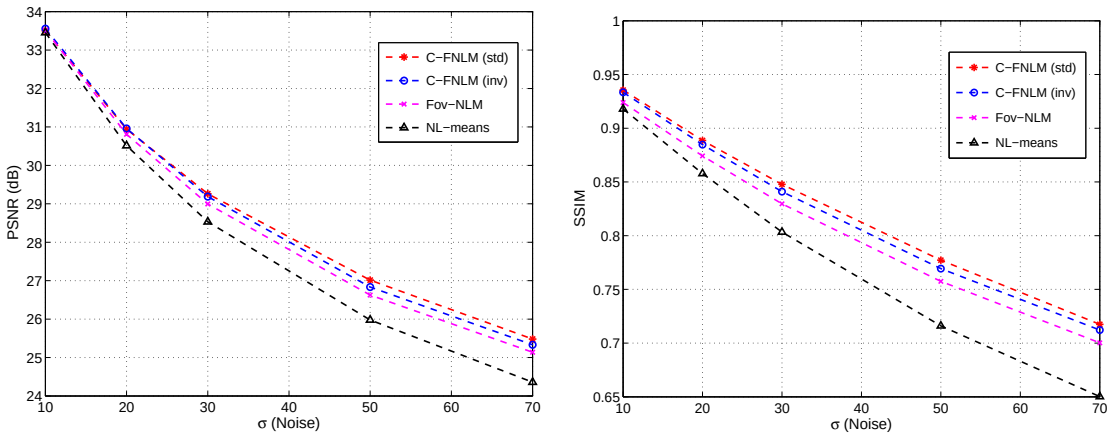
(d) Unified Output Image, with Standard CM. (e) Unified Output Image, with Inside-Out CM. (f) Unified Output Image, with Uniform CM.

Figure 5.2: Top: Illustration of a color-mixed foveated patch extracted from noisy *Lena* ($\sigma = 30$) and having size 11×11. It must be noted that the C-FNLM algorithm operates in a patch-wise non-local manner within a search window, and for each color-mixing array shown in (a) Standard (refer Fig. 5.1c), (b) Inside-Out, and (c) Uniform, we display the corresponding output patches. Color-mixed foveation preserves the original image structures better than windowing. Bottom: For visualization purposes, we display the Unified Outputs for various color-mixing arrays of size 301×301.

Lena *Barbara* *Boats* *Hill*

Figure 5.3: The four 512×512 color images y used in denoising experiments.

It is imperative for the reader to note that the foveal color vision is formalized by the unified output image (Fig. 5.2d) using the standard color mixing array. This interpretation of color vision stems from the notion that cones act as the color sensors of the HVS and are responsible for trichromatic vision, whereas rods contribute monochromatic color quality [66], [39]. It should also be noted that high frequencies enjoy a shorter-range correlation than the low frequencies, and since color-mixed foveation attenuates the high frequency content at the periphery of the patch, it indeed emphasizes the information useful for the purpose of non-local denoising of the patch center. The denoising result for the unified operator with standard and inside-out (Fig. 5.2e) CM is presented in Fig. 5.4.



(a) PSNR (dB) Comparision (b) SSIM Comparision

Figure 5.4: Scatterplots of PSNR (dB) and SSIM of the standard NL-means vs Foveated NL-means vs C-FNLM, for two combinations of color-mixing array - the inside-out and standard CM array. Each point represents the PSNR value (or, SSIM score) achieved for the best parameter combination of patch size and search neighbourhood, determined from Fig. 5.5, 5.6, averaged over the test images in Fig. 5.3, at a given noise level. When $\sigma \geqslant 30$, Foveated NL-means and C-FNLM outperforms the standard NL-means in all considered configurations; while when $\sigma = 10$, the best setting for all the three methods give approximately the same results. The two CM variants yield nearly the same performance, with negligible differences in favor of the standard CM.

As a reference code of NL-means and Foveated NL-means, we use the original MATLAB code by [43] and [31], respectively, and apply them to each color channel separately. To enable a fair comparision between the algorithms, where $d_{\mathrm{j}}^{\mathrm{FOV}}$ is obtained from a color foveation operator $\mathcal{F}_{\mathrm{RGB}}$ (Sec. 5.1), we test the algorithms with several configurations of *patch size* (ranging from 5×5 to 19×19) and *search neighbourhood* (ranging from 7×7 to 41×41), while the tuning parameter h is set equal to σ as this choice is found to yield the optimal results for standard NL-means, Foveated NL-means, as well as for the C-FNLM, and is consistent with the compatibility requirement. $\mathcal{F}_{\mathrm{RGB}}$ implements the isotropic foveation operator $\mathcal{F}$ which is constructed from the windowing kernel $\mathbf{k}$ and is identical over all the color channels. Therefore, the algorithms differ only in the employed patch distance.

The parameters which maximize denoising performance, for $\sigma = 10, 20, 30, 50, 70$ respectively, are: for C-FNLM and Foveated NL-means, the patch size is set to $(5\times5), (9\times9), (11\times11), (17\times17), (19\times19)$ and the search neighbourhood is (13×13), irrespective of σ ; for NL-means, the optimum results are obtained for a patch size of (5×5) and a search neighbourhood of (11×11), for all σ.

Figures 5.5, 5.6 show that the C-FNLM and Foveated NL-means substantially outperform (by about 1.1 dB PSNR and 0.3 SSIM units) the standard NL-means, especially under heavy noise ($\sigma \geqslant 30$), for all the considered configurations.

At low noise levels ($\sigma = 10$) all three methods perform best when using small patches, and give comparable results. As the noise level increments ($20 < \sigma < 90$) the foveation-based methods show a numerical improvement as the patch size increases. The SSIM scores too indicate a favorability for an increasing patch size corresponding to the increase in noise levels.

To further assess the performance of our algorithm in terms of PSNR, we compared the results with the patch-wise implementation of NL-means given in IPOL [10], where the parameters have been fixed to achieve the maximum gain in PSNR value. It was found that under heavy noise ($\sigma \geqslant 30$), C-FNLM outperforms the IPOL's version of NL-means (by about 0.3 dB). As a figure of merit, when denoising *Lena* corrupted with noise $\sigma = 30$, C-FNLM with standard color-mixing achieves a PSNR of 30.58 dB, versus 30.21 dB for patch-wise NL-means. Similarly, when denoising *Boats* corrupted with noise $\sigma = 70$, C-FNLM achieves a PSNR of 25.38 dB, versus 25.15 dB for patch-wise NL-means. The output of the proposed algorithm is characterized by better contrast, and increased detail preservation. The visual difference between the outputs is shown in Fig. 5.7.

5.5 Conclusions

In this thesis, we introduce and test the efficiency of a unified color-mixing foveation operator which exploits the correlation between color channels in an image. This efficacy become apparent when all variants of NL-means for color images, including the patch-wise NL-means [10], are outperformed by C-FNLM and Foveated NL-means. The performance gap between C-FNLM and Foveated NL-means is less substantial than the improvement achieved by introducing isotropic foveation in the windowing based NL-means. Nevertheless, such improvement is particularly meaningful as it highlights that the color-mixed foveation yields a stronger prior than the windowing conventionally used in NL-means for measuring non-local similarity in color image denoising.

We have constructed an isotropic family of color-mixing foveation operators, which simultaneously performs the foveation and color-mixing operations. The designs of the PSFs have installed an explicit connection between traditional windowed self-similarity and the color-mixed foveated self-similarity.

Table 5.1 shows that the overall computation time of C-FNLM is marginally higher than that of Foveated NL-means, which in-turn is marginally higher than NL-means. This overhead is due to the computation of the color-mixed foveated patches (in lieu of either foveation or windowing), which is executed only once on the whole image, before computing $d_{\mathrm{j}}^{\mathrm{FOV}}$, where $\mathrm{j} \in \{\mathrm{R}, \mathrm{G}, \mathrm{B}\}$. Once the patches are computed, the time needed for pairwise patch comparisions has approximately identical complexity and is the most time consuming operation in the algorithm.

Computation Time (sec.) for MATLAB single-thread on Intel i7-5500U @ 2.4GHz.

Patch Size: 5×5, Search nbd: 11×11			Patch Size: 11×11, Search nbd: 13×13		
NL-Means	Fov. NLM	C-FNLM	NL-Means	Fov. NLM	C-FNLM
2.1 s	5.2 s	3.8 s	7.3 s	17.7 s	22.4 s
+	+	+	+	+	+
66.5 s	67.2 s	68.5 s	331.9 s	333.4 s	334.6 s

Table 5.1: Computation time of NL-means vs Foveated NLM vs C-FNLM (for $\sigma = 30$). The patch size and search neighbourhood of the first three columns are those that yield best overall results for NL-means, while those of the last three columns are optimal for Fov. NLM and C-FNLM, as in Fig. 5.5, 5.6. For each algorithm, we report the average computation time in seconds over the images in Fig. 5.3 and three noise realizations, separating the time needed for either windowing, foveating or color-mixed foveating the patches (top) from that needed for computing weights and averaging (bottom).

Finally, we remark that despite the improvement achieved by introducing the color-mixing paradigm in Foveated NL-means, the performance of both C-FNLM and Foveated NL-means is still inferior to sophisticated non-local filters, e.g. C-BM3D [21] or NL-Bayes [12]. As a figure of merit, when denoising *Lena* corrupted with noise $\sigma = 30$, C-FNLM with standard color-mixing achieves a PSNR of 30.58 dB, while C-BM3D achieves 31.59 dB and NL-Bayes achieves 31.39 dB. However, our contribution is not intended to be the development of a novel denoising algorithm, but rather the exploration of a new form of non-local self-similarity for color images which is consistent with HVS features.

5.5.1 Additional Remarks

It is possible to develop a framework for "joint denoising and demosaicing" of noisy color filter array (CFA) data by utilizing the cross-channel paradigm with Foveated NL-means. The two-stage algorithm entails directly applying the Color-mixed Foveated NL-means (C-FNLM) in the CFA domain during the first iteration, to exploit the non-local similarity, despite the absence of local-smoothness in the underlying mosaic structure. We also reformulate the weighing scheme so as to account for correlated, non-i.i.d. noise, principally during the second iteration of the proposed algorithm.

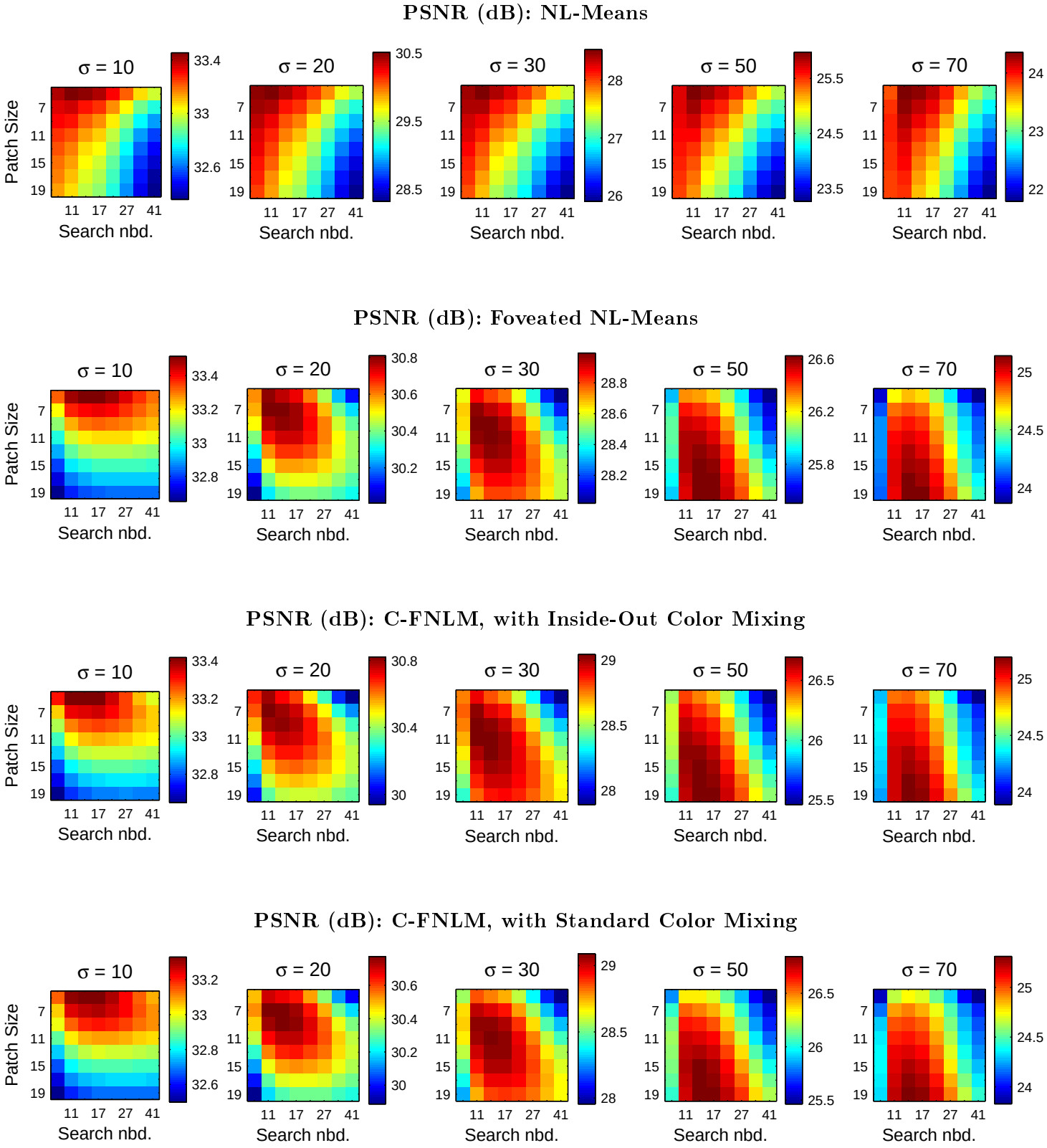


Figure 5.5: Performance of the standard NL-means, Foveated NL-means, and C-FNLM, in terms of PSNR (dB), while varying the search radius and patch size. The NL-means and Foveated NL-means results are obtained by filtering the color channels separately. The denoising values is averaged over the four test images in Fig. 5.3, each corrupted by 3 different noise realizations.

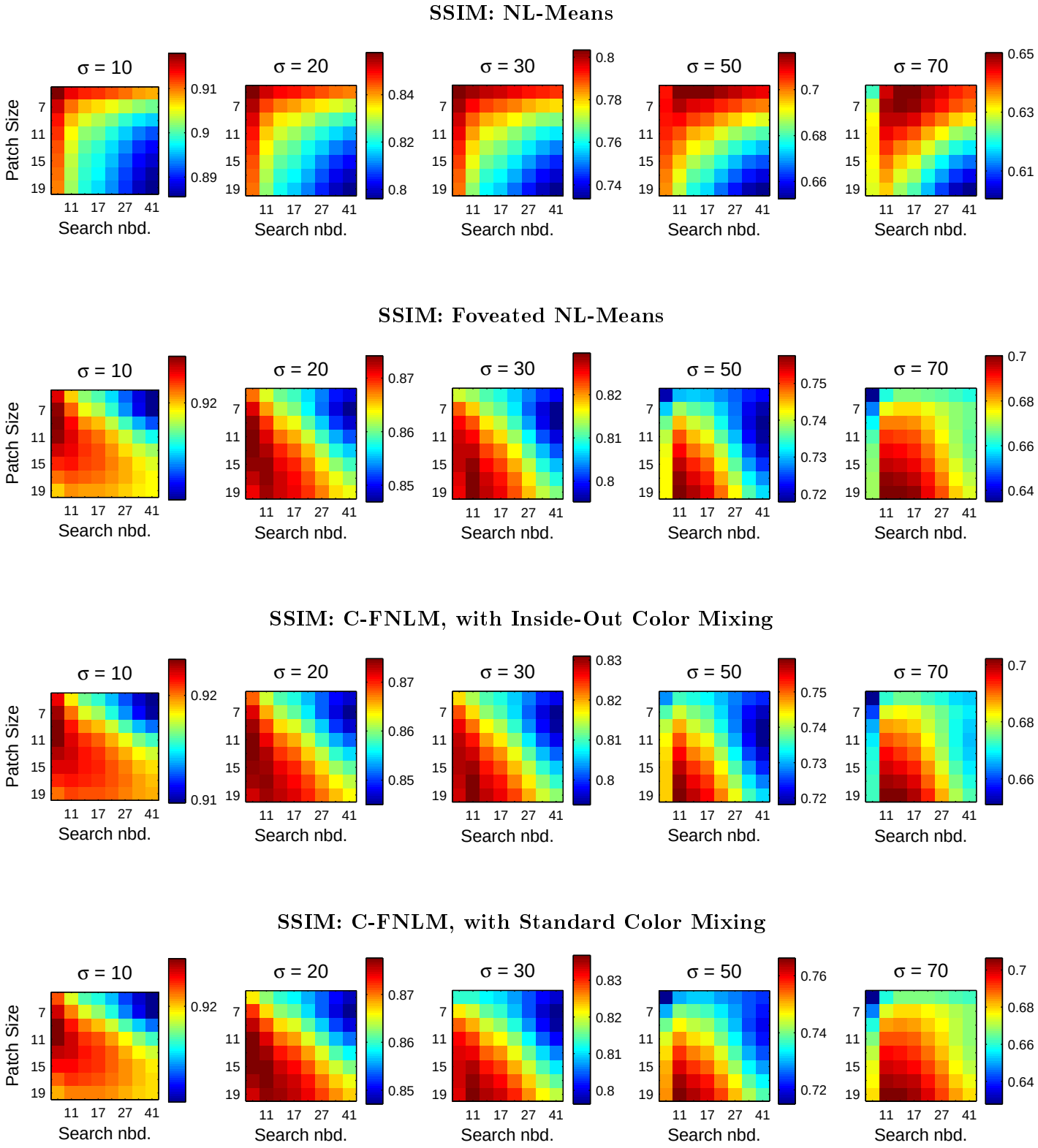


Figure 5.6: Performance of the standard NL-means, Foveated NL-means, and C-FNLM, in terms of SSIM score, while varying the search radius and patch size. The NL-means and Foveated NL-means results are obtained by filtering the color channels separately. The denoising values is averaged over the four test images in Fig. 5.3, each corrupted by 3 different noise realizations.

Patch Size $= 5 \times 5$ **Search Neighbourhood** $= 11 \times 11$

Hill Original | *NL-Means* (29.25, 0.775) | *Fov. NL-Means* (29.54, 0.789) | *C-FNLM* (29.50, 0.787)

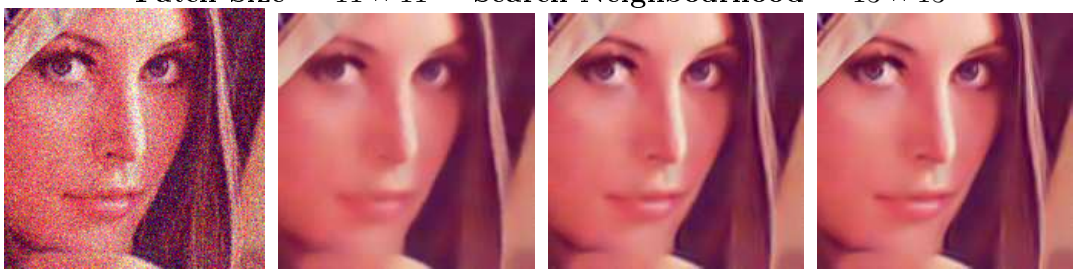
Patch Size $= 9 \times 9$ **Search Neighbourhood** $= 13 \times 13$

Hill, noisy ($\sigma = 20$) (22.11, 0.597) | *NL-Means* (28.69, 0.745) | *Fov. NL-Means* (29.47, 0.790) | *C-FNLM* (29.51, 0.791)

Patch Size $= 5 \times 5$ **Search Neighbourhood** $= 11 \times 11$

Lena Original | *NL-Means* (29.96, 0.832) | *Fov. NL-Means* (30.30, 0.842) | *C-FNLM* (30.35, 0.843)

Patch Size $= 11 \times 11$ **Search Neighbourhood** $= 13 \times 13$



Lena, noisy ($\sigma = 30$) (18.59, 0.343) | *NL-Means* (29.46, 0.821) | *Fov. NL-Means* (30.55, 0.854) | *C-FNLM* (30.58, 0.855)

Figure 5.7: Comparision between outputs of the NL-means algorithm, the FNLM and the proposed C-FNLM. The numbers between parentheses are the PSNR (dB) and SSIM scores computed for the entire image, not just the displayed fragment of size 175×175 pixels. Results are given under two combinations of patch size and search neighbourhood, one ideal for FNLM and C-FNLM, another for NL-means (see Fig. 5.5, 5.6). The standard CM array is used while implementing C-FNLM for the images.

Figure 5.8: Comparision between outputs of the NL-means algorithm, the FNLM and the proposed C-FNLM. The numbers between parentheses are the PSNR (dB) and SSIM scores computed for the entire image, not just the displayed fragment of size 175 × 175 pixels. Results are given under two combinations of patch size and search neighbourhood, one ideal for FNLM and C-FNLM, another for NL-means (see Fig. 5.5, 5.6). The standard CM array is used while implementing C-FNLM for the images.

BIBLIOGRAPHY

[1] A. Alahi, R. Ortiz, and P. Vandergheynst. FREAK: Fast Retina Keypoint. *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

[2] A. Alahi, P. Vandergheynst, M. Bierlaire, and M. Kunt. Cascade of Descriptors to Detect and Track Objects across any Network of Cameras. *Computer Vision and Image Understanding*, 2011.

[3] J. Allen. Short term Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1977.

[4] G. Boracchi. Foveated Self-Similarity in Nonlocal Image Filtering. Technical report, Politecnico di Milano, 2013.

[5] A. Bovik. *Handbook of Image and Video Processing.* Academic Press, 2000.

[6] A. Buades, B. Coll, and J. Morel. A Non-local Algorithm for Image Denoising. *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

[7] A. Buades, B. Coll, and J. Morel. A review of Image Denoising Algorithms, with a new one. *Multiscale Modeling Simulation*, 2005.

[8] A. Buades, B. Coll, and J. Morel. Non-local Image and Movie Denoising. *International Journal of Computer Vision*, 2008.

[9] A. Buades, B. Coll, and J. Morel. Image Denoising Methods : A New Nonlocal Principle. *SIAM Review*, 2010.

[10] A. Buades, B. Coll, and J. Morel. Non-Local Means Denoising, *Image Processing On Line*. `http://www.ipol.im/pub/art/2011/bcm_nlm`, 2011.

[11] A. Buades, B. Coll, J. Morel, and C. Sbert. Non local demosaicing, 2007.

[12] A. Buades, M. Lebrun, and J.M. Morel. A Non-local Bayesian Image Denoising Algorithm. *SIAM Journal on Imaging Sciences*, 2013.

[13] Harold Christopher Burger. *Modelling and Learning Approaches to Image Denoising.* PhD thesis, Eberhard Karls Universität Tübingen, 2012.

[14] E. Candes, Y. Eldar, D. Needell, and P. Randal. Compressed Sensing with Coherent and Redundant Dictionaries. *Applied and Computational Harmonic Analysis*, 2011.

[15] L.M. Chalupa and J.S. Werner. *The Visual Neurosciences.* MIT Press, 2004.

[16] P. Chatterjee and P. Milanfar. Is Denoising Dead ? *IEEE Transactions on Image Processing*, 2010.

[17] C.M. Cicerone. Color Appearance and the Cone Mosaic in Trichromacy and Dichromacy. *Color Vision Deficiencies*, 1990.

[18] C. Curcio, K. Sloan, R. Kalina, and A. Hendrickson. Human Photoreceptor Topography. *Journal of Comparative Neurology*, 1990.

[19] K. Dabov, A. Foi, and K. Egiazarian. Video Denoising by Sparse 3D Transform-Domain Collaborative Filtering. *European Signal Processing Conference*, 2007.

[20] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image Denoising with Block-Matching and 3D Filtering. *SPIE International Society for Optical Engineering*, 2006.

[21] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Color Image Denoising via Sparse 3D Collaborative Filtering with Grouping Constraint in Luminance-Chrominance Space. *IEEE International Conference on Image Processing*, 2007.

[22] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image Denoising by Sparse 3D Transform-Domain Collaborative Filtering. *IEEE Transactions on Image Processing*, 2007.

[23] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image Restoration by Sparse 3D Transform-Domain Collaborative Filtering. Technical report, Tampere University of Technology, 2008.

[24] A. Duchowski. *Eye Tracking Methodology.* Springer-Verlag, 2007.

[25] R.O. Duncan and G.M. Boynton. Cortical Magnification within Human Primary Visual Cortex Correlates with Acuity Thresholds. *Neuron*, 2003.

[26] V. Duval, J. Aujol, and Y. Gousseau. A Bias-Variance Approach for the Nonlocal Means. *SIAM J. Imaging Sciences*, 2011.

[27] A. Efros and T. Leung. Texture Synthesis by Non Parametric Sampling. *IEEE International Conference of Computer Vision*, 1999.

[28] C. Enroth-Cugell and J.G. Robson. The Contrast Sensitivity of Retinal Ganglion Cells of the Cat. *Journal of Physiology*, 1966.

[29] R. Fattal, M. Agrawala, and S. Rusinkiewicz. Multiscale Shape and Detail Enhancement from Multi-light Image Collections. *ACM Transactions on Graphics*, 2007.

[30] A. Foi. Patch Foveation in Nonlocal Imaging. Technical report, Tampere University of Technology, 2012.

[31] A. Foi and G. Boracchi. Anisotropic Foveated NL-means filter, MATLAB Code. `http://www.cs.tut.fi/~foi/FoveatedNL/index.html`, 2012.

[32] A. Foi and G. Boracchi. Foveated Self-Similarity in Nonlocal Image Filtering. *SPIE Conference on Human Vision and Electronic Imaging*, 2012.

[33] A. Foi and G. Boracchi. Anisotropically Foveated Nonlocal Image Denoising. *IEEE International Conference on Image Processing*, 2013.

[34] A. Foi and G. Boracchi. Anisotropically Foveated Self-Similarity. *Signal Processing with Adaptive Sparse Structured Representations*, 2013.

[35] A. Foi and G. Boracchi. Foveated Nonlocal Self-Similarity. *Submitted*, 2014.

[36] A. Foi and G. Boracchi. Nonlocal Foveated Principal Components. *IEEE Workshop on Statistical Signal Processing*, 2014.

[37] J. Freeman and E.P. Simoncelli. Metamers of the Ventral Stream. *Nature Neuroscience*, 2011.

[38] A. Gillespie, A. Kahle, and R. Walker. Color Enhancement of Highly Correlated Images. *Remote Sensing of Environment*, 1987.

[39] J. Gordon and I. Abramov. Color Vision in the Peripheral Retina. II. Hue and Saturation. *Journal of the Optical Society of America*, 1977.

[40] G.Petschnigg, R. Szeliski, and M. Agrawala. Digital Photography with Flash and No-Flash Image Pairs. *ACM Transactions on Graphics (TOG) - SIGGRAPH*, 2004.

[41] R.L. Gregory. *Eye and Brain: The Psychology of Seeing.* Princeton University Press, 1990.

[42] E. Hecht. *Optics.* Addison Wesley, 1987.

[43] J. Herrera and A. Buades. Non-Local Means Filter, MATLAB Code. `http://www.mathworks.com/matlabcentral/fileexchange/13176-non-local-means-filter`, 2008.

[44] M. Hogan and J. Weddell. Histology of Human Eye: an Atlas., 1971.

[45] A. Jacquin. Image Coding based on a Fractal Theory of Iterated Contractive Image Transformations. *IEEE Transactions on Image Processing*, 1992.

[46] J.A.M. Jennings and W. Charman. Analytic Approximation of the Off-axis Modulation Transfer Function of the Eye. *Vision Research*, 1997.

[47] C. Joselevitch. Human Retinal Circuitry and Physiology. *Psychology & Neuroscience*, 2008.

[48] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola. From Local Kernel to Nonlocal Multiple-model Image Denoising. *International Journal of Computer Vision*, 2010.

[49] C. Kervrann and J. Boulanger. Optimal Spatial Adaptation for Patch-based Image Denoising. *IEEE Transactions on Image Processing*, 2006.

[50] C. Knaus and M. Zwicker. Dual-Domain Image Denoising. *IEEE International Conference on Image Processing*, 2013.

[51] E. Kowler. Eye Movements: The Past 25 Years. *Vision Research*, 2011.

[52] S.W. Kuffler. Discharge Patterns and Functional Organization of Mammalian Retina. *Journal of Neurophysiology*, 1953.

[53] S. Leutenegger, M. Chli, and R. Siegwart. Coarse–to–fine Face Detection. *International Journal of Computer Vision*, 2011.

[54] C. Louchet and L. Moisan. Total Variation as a Local Filter. *SIAM J. Imaging Sciences*, 2011.

[55] M. Mäkitalo. *Exact Unbiased Inverse of the Anscombe Transformation and its Poisson-Gaussian Generalization.* PhD thesis, Tampere University of Technology, 2013.

[56] M. Mäkitalo and A. Foi. Optimal Inversion of the Anscombe Transformation in Low-count Poisson Image Denoising. *IEEE Transactions on Image Processing*, 2011.

[57] M. Mäkitalo and A. Foi. Optimal Inversion of the generalized Anscombe Transformation for Poisson-Gaussian Noise. *IEEE Transactions on Image Processing*, 2013.

[58] J.C. Maxwell. On Color Vision. *Proceedings of the Royal Institute, Great Britain*, 1872.

[59] J. Moreland and A. Cruz. Color Perception with the Peripheral Retina. *Optica Acta 6*, 1959.

[60] K.T. Mullen and F.A. Kingdom. Differential Distributions of Red-green and Bue-yellow Cone Opponency across the Visual Field. *Visual Neuroscience*, 2002.

[61] S. Postec. *Quelques remarques en débruitage des Images liées à des Propriétés de Similarité, de Régularité et de Parcimonie.* PhD thesis, Université de Bretagne-Sud, 2012.

[62] R.W. Rodieck. Quantitative Analysis of Cat Retinal Ganglion Cell Response to Visual Stimuli. *Vision Research*, 1965.

[63] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. ORB: An Efficient Alternative to SIFT or SURF. *IEEE Conference on Computer Vision*, 2011.

[64] Y. Sasaki, R. Rajimehr, B. Kim, and *et al.* The Radial Orientation Effect in Human and Non-human Primates. *Journal of Vision*, 2006.

[65] C. Tomasi and R. Manduchi. Bilateral Filtering for Gray and Color Images. *International Conference on Computer Vision*, 1998.

[66] C. Vakrou, D. Whitaker, and *et al.* Functional Evidence for Cone-specific Connectivity in the Human Retina. *Journal of Physiology*, 2005.

[67] D. Van De Ville and M. Kocher. SURE-Based Non-Local Means. *IEEE Signal Processing Letters*, 2009.

[68] B. Wandell. *Foundations of Vision.* Sinauer Associates, 1995.

[69] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image Quality Assessment: From Error Measurement to Structural Similarity. *IEEE Transactions on Image Processing*, 2004.

[70] A. Wohrer. The Vertebrate Retina: A Functional Review., 2008.

[71] W. D. Wright. *Researches on Normal and Defective Colour Vision.* Oxford, England, 1946.

[72] G. Yu and G. Sapiro. DCT Image Denoising: a simple and effective Image Denoising Algorithm. *Image Processing On-Line*, 2011.

[73] D. Zhang and Z. Wang. Image Information Restoration Based on Long-Range Correlation. *IEEE Trans. on Circuits and Systems on Video Technology*, 2002.